
17

PLATAFORMAS DE DATOS ABIERTOS: DISPONIBILIDAD DE MERCADO

*Viena Muirragui Irrazábal, Fernando Pacheco Olea, Edwin León
Plúas, Fabricio Guevara Viejó.*

*Recibido: agosto de 2016
Aprobado: diciembre de 2016*

PLATAFORMAS DE DATOS ABIERTOS: DISPONIBILIDAD DE MERCADO OPEN DATA PLATFORMS: MARKET AVAILABILITY

Viena Muirragui Irrazábal, Fernando Pacheco Olea, Edwin León Plúas, Fabricio Guevara Viejó

Universidad Estatal de Milagro
vmuirragui1@unemi.edu.ec

Resumen

El documento presenta las oportunidades que dentro del contexto de disponibilidad de mercado existen para el uso de las plataformas open data, ya que no todas las organizaciones tienen recursos propios que sean suficientes para implantar sus plataformas desde una base inicial, siendo por lo tanto necesario requerir aquellas que han sido desarrolladas por empresas expertas en ello; si se considera tomar ésta opción tendrían a resolver el proceso, sin embargo podría constituirse en una pérdida de capacidad de llegar a personalizar y adecuar integralmente la plataforma a las necesidades específicas de la organización que las requiera. Tomando en consideración lo expuesto, en este artículo de revisión se detallan características con las que cuentan algunas de las plataformas consideradas principales por su aplicabilidad que están disponibles en el mercado, sea a través de un modelo de software gratuito o libre y, también, con licencia comercial, en otros casos. Los datos que han sido considerados para este análisis se los referencia con respecto a la información que los propios proveedores han proporcionado por medio de documentación que se encuentra publicada por ellos en la web, como también se ha considerado la experiencia con versiones trial y ejemplos utilizables de manera abierta.

Palabras clave: *disponibilidad, plataformas de datos abiertos, modelo de software, uso comercial, documentación web*

Abstract

The paper presents the opportunities within the context of market availability exist for the use of open data platforms, since not all organizations have their own resources which are sufficient to implement their platforms from an initial base, being therefore necessary to require those that have been developed by expert companies in it; considering taking this option would solve the process, however it could become a loss of ability to reach fully customize and adapt the platform to the specific needs of the organization that requires them. Considering the above, in this review article features are detailed with that feature some of the platforms major consideration for its applicability that are available in the market, either through a model of free and open source software and also licensed trade in other cases. The data that have been considered for this analysis is the reference regarding the information that the suppliers have provided through documentation that is published by them on the web, as has also been considered the experience with trial versions and examples used openly.

Keywords: *availability, open data platforms, software model, commercial use, web documentation*

*Recibido: agosto de 2016
Aprobado: diciembre de 2016*

1.Introducción

El intercambio y la preservación de datos es un tema de interés a escala internacional para quienes en determinados momentos financian, gestionan y generan la investigación, lo que desde 2003 recogen aspectos relacionados al acceso abierto a la información a través del *National Institutes of Health* (NIH). Borrego (2012), toma como referencia el informe *opportunities for data exchange* (ODE, 2011), presenta un 60% el número de investigadores que estarían dispuestos a utilizar datos primarios de otros agentes y, un 40% muestra poca disposición para compartir sus aportes. La investigación científica como tal genera mayor volumen de datos digitales.

Cabe por lo consiguiente destacar que, si bien es cierto éstas plataformas procuran cierto grado de flexibilidad y, en ocasiones algunos componentes de extensión, su nivel de personalización es limitado, requiriendo por lo tanto también mayor esfuerzo en su funcionalidad. Sin embargo, tal y como se presenta a continuación, gran parte de las características que toda plataforma *open data* tiene, busca que se ponga especial atención al nivel de compatibilidad que los estándares existentes ofrecen a nivel internacional.

1. CKAN Data management system

Un sistema de gestión de datos de gran alcance es sin duda *CKAN*ⁱⁱ el cual permite que los datos sean de fácil acceso cuya herramienta permite hacer que la publicación sea clara, así también permite compartir información, realizar búsquedas y hacer uso de datos. *CKAN* está orientado a editores de datos, sean estos de gobiernos nacionales, regionales, seccionales a más de empresas y organizaciones sin fines de lucro que requieren sus datos estén disponibles y sean abiertos; es también una herramienta creada para promover la publicación de catálogos de datos y es auspiciada por la asociación *Open Knowledge*ⁱⁱⁱ. Es posible de igual forma compararlo con un sistema de gestión de contenido tradicional (*CMS*)^{iv}, con la particularidad que está orientado a la publicación de datos y no de otros contenidos cuya referencia está dada en crear, editar, gestionar y publicar contenido digital multimedia en diferentes tipos de formatos. *CKAN* permitirá publicar y administrar conjuntos de colecciones de datos utilizando pocos recursos y, cuando éstos sean publicados, los usuarios podrán revisarlos a más de pre-visualizarlos utilizando mapas, gráficos y tablas. (Peset, et al., 2011).

La plataforma posee las siguientes características:

1.1 Publicación y gestión de datos

Los datos son cargados utilizando la importación directa de archivos, empleando interfaces de programación e inclusive aplicando una interfaz *web* sencilla, la recolección de los datos se la realiza de forma automatizada de otros repositorios; brinda la posibilidad de albergar los datos en la propia plataforma con la respectiva gestión de categorías y permisos. (Ortega y Aguillo, 2006, 270-277)

1.2 Búsqueda y reutilización

Afirman Martínez y Rodríguez (2003) “que la búsqueda debe ser capaz de presentar resultados relevantes combinando búsquedas simples con otros mecanismos más avanzados como las búsquedas facetadas a través de filtros compuestos”. Siendo por lo consiguiente que las búsquedas sintácticas realizadas considerando palabras clave o filtradas en algunos casos, considerando en otros también búsquedas por facetas en base a los metadatos que se proporcionan ayudan a agilizar los procesos.

1.3 Metadatos

Los metadatos, según manifiesta (Codina, 2003, 149-152) son “información sobre la información y son, en realidad, una antigua fórmula”; tal es así que los catálogos de las bibliotecas, a modo de ejemplo, son metadatos. Por lo que éste amplio conjunto de metadatos que se encuentran disponibles, incluyen también metadatos geospaciales conforme la directiva *INSPIRE (Infrastructure for Spatial Information in Europe)*, otorgando la posibilidad de incorporar metadatos nuevos y personalizados.

1.4 Herramientas de aporte a la comunidad

Los proyectos que soportan la apertura de datos llegan a convertirse en ejemplos de proyectos para lograr gestionar el cambio necesario con respecto a la visualización, gestión y localización de la información. Convirtiéndose en un elemento fundamental para lograr el éxito, contando con un amplio soporte de *reutilizadores* y usuarios de los datos. (Manovich, 2013). Logrando tener la posibilidad de compartir diversos conjuntos de datos en redes sociales, con suscripciones a contenidos y, por ende actualizaciones tomando como punto de referencia las herramientas para recoger el *feedback* de la comunidad”.

1.5 Visualización

La visualización de la información constituye un factor primordial para que las plataformas *open data* tengan aspectos amigables, cercanos, comprensibles y accesibles a los datos. Siendo de suma importancia incorporar ciertas capacidades básicas de geolocalización y comparación de los datos ya existentes en los catálogos de datos tal como lo expresa (Sancho, 2014) dentro del contexto de aproximación a la visualización de datos. La posibilidad de visualizar los datos que están debidamente estructurados considerando tablas dinámicas o gráficos básicos se convierten en espacios de difusión de información; donde la geolocalización de los cuentan con componentes geográficos que les permita ser ubicados a tiempo.

Adicionalmente, *CKAN* admite una personalización avanzada, tanto en la apariencia del catálogo al igual que en su funcionalidad, utilizando sus mecanismos esenciales para el desarrollo de temas y extensiones.

Existe la versión online *demo* de *CKAN* utilizada para pruebas, la que está disponible de forma *free* como *software* libre, así también sus creadores brindan servicios adicionales de desarrollar, alojar, dar soporte y formación contactando previamente para obtener un coste del producto. *CKAN* es empleado por varias organizaciones a nivel internacional.

2. OGoov – Plataforma de tipo Gobierno Abierto

Esta plataforma de gobierno abierto, desarrollada por *Viavansi* es otro ejemplo de las plataformas ya elaboradas y su comercialización es más frecuentemente utilizada en España a través de la empresa de telecomunicaciones Telefónica, su particularidad reside en brindar una serie de funciones que logran ser combinables entre sí según sea la orientación que los gobiernos con acceso abierto requieran dar mayor prioridad: de donde se coligen que estos sean especialmente datos abiertos, con transparencia y, sobretodo cuenten con participación ciudadana.

OGoov cuyo link oficial es <http://www.ogoov.com> se configura para ofrecer de forma indistinta una plataforma *open data*, una plataforma de transparencia usando todos los módulos que ella ofrece.

La arquitectura, como expresa (Ronda, 2008) “muestra cómo las categorías de información se relacionan con los procesos de

negocio y cómo deben estar conectadas para facilitar la toma de decisiones”, por lo que la solución para esta plataforma *OGoov* al usar este tipo de *software* se basa en dos piezas clave, el gestor *open data* *CKAN*, y el portal como tal, desarrollado sobre WordPress^{vi} con la finalidad de enaltecerlas capacidades fundamentales que *CKAN* dispone. Así también, ofrece la posibilidad de incorporar un *middleware*^{vii} *Real Time Open Data (RTOD)* que logre dotar de forma dinámica los conjuntos de datos en nueve formatos distintos, adicionando soporte para *Linked Open Data* (Agenjo y Hernández, 2010) los que son: datos contextualizados y datos enriquecidos semánticamente, en algunos casos desde las bibliotecas digitales a otros datos, desde datos no bibliográficos a las bibliotecas digitales utilizando un servidor semántico perfeccionado con Apache Marmotta^{viii} y contando adicionalmente con un punto de consulta *SPARQL*.^{ix}

Sus características son:

1. Soporte de cumplimiento de datos abiertos (*open data*)
2. Unión automática de contenidos con los catálogos de datos
3. Gestionar datos abiertos de forma eficiente con conexión en “tiempo real” con las fuentes de datos.
4. Posee herramientas capaces de procesar datos en tablas, generando gráficos y mapas.
5. Su instalación es propia o a través del servicio denominado “*cloud*”^x
6. Es accesible en varios idiomas

3. Socrata Open Data Portal

Es una plataforma escalable de donde colige (Cano-Parra, 2014) que constituyen pequeñas partes reutilizables de medios instructivos que orientan la publicación de datos en la nube, facilitando la creación de proyectos *open data* sostenibles, otorgando un conjunto basto de funcionalidades, destacándose:

3.1 Publicación y gestión de datos

Conseguir usuarios, captar su interés y lograr que participen del mensaje compartido resulta primordial para lograr respuesta y colaboración. (Cebrián, 2009) indica que utilizar herramientas elementales de difusión integradas en las plataformas de publicación, ayudarán a cautivar la atención desde el inicio para así difundir el mensaje de forma rápida y en mayor proporción.

Socrata, anexa herramientas cuya función principal es la de obtener datos, como hojas de

cálculo, sistemas de archivo, bases de datos transaccionales y, bases de datos en tiempo real, en ella se pueden realizar actualizaciones automáticas, cuenta con una interfaz de programación amigable y guías para el flujo de trabajo. La gestión de los metadatos es flexible contando con el soporte estándar DCAT ^{xi} permitiendo la utilización de diversos tipos de metadatos y, vocabularios personalizados.

3.2 Acceso y visualización

Las características de búsqueda avanzada a través de palabras clave, etiquetas y metadatos como expresan (Ferrer et al., 2011) permiten que estas tengan la posibilidad de aplicar filtros, agruparlos y ordenarlos para que su búsqueda sea mas rápida y clara; de donde las búsqueda por fases sean visualizadas utilizando variedad de gráficos, al igual que mapas que consienten aplicar filtros de búsquedas avanzados con respecto a los datos.

3.3 Creación a partir de datos

Indica (Montuschi, 2001) que referirse al conocimiento, este existe en cuanto tiene alguna aplicación en acciones concretas de difusión de información de los datos con los que se cuentan, tal es el caso que cuando se trata de acceso a los datos en los diversos formatos que se presentan en la web otorgan la posibilidad de incorporar fácilmente los conjuntos de datos en los contenidos de otros portales que les sirvan de canales de difusión; la interfaz de programación facilita el desarrollo de las aplicaciones con los datos basándose en estándares OData (*Open Data Protocol*).

4. Junar – The Open Data Platform

Corresponde ahora analizar a *Junar*, es una plataforma *open data* en la nube (Aguilar, 2013) facilita que los datos publicados por gobiernos, empresas u organizaciones tengan visibilidad y accesibilidad para los usuarios de forma práctica y segura. Con Junar se pueden escoger los datos que se vayan a publicar, al igual que el tiempo y la forma cómo van a ser presentados al público. Los servicios que ella ofrece son:

4.1 Recolección de datos

La plataforma como tal, es un medio que permite recopilar todos los datos a compartirse a través de un punto común; estos datos pueden ser introducidos utilizando archivos individuales o en la *web* a través de enlaces (*link*).

4.2 Mejora de datos

Junar permite convertir datos en información de forma productiva utilizando diversos tipos de distintas visualización, tablas de datos y mapas. Las visualizaciones son posibles de organizarse elaborando cuadros de mando (*Balanced Score Card*) personalizados haciendo seguimiento de los más utilizadas. (Ballvé y Salas, 2006).

4.3 Publicación

En Junar, el proceso de publicación toma en consideración los procesos al interior de la organización, logrando asignar los permisos suficientes para cada usuario cuando éste pretenda la recopilación, mejora y publicación de los datos. Este método adiciona la gestión de datos netamente privados, del cual su uso es de exclusividad interna; así también, ofrece una interfaz de programación (*API*) totalmente integrada para ayuda de los desarrolladores, tanto de servicios como los de aplicaciones de forma que puedan reutilizar los datos de manera práctica, segura y rápida.

4.4 Uso.- análisis

Es posible realizarse su seguimiento en tiempo real, ya que cuenta con estadísticas proporcionadas por sus desarrolladores con respecto a la utilización de las distintas formas de visualizar su interfaz, posee paneles de control API, otros aspecto de gran interés, es que dicha plataforma puede integrar su funcionalidad con *Google Analytics*^{xiii} añadiendo estadísticas de uso y seguimiento de otros sitios donde éste se esté aplicando.

Otra experiencia que Junar permite instaurar, es el de las conversaciones a través de redes sociales con los usuarios que acceden a los datos, teniendo como finalidad entender cuáles son sus requerimientos al momento que los datos sean publicados y ellos, los usuarios, tengan acceso a suscribirse.

Organizaciones que actualmente utilizan Junar son países como: Chile^{xiii} y Costa Rica, en las ciudades de San José en California y Lima en Perú, en España la Diputación de Cádiz, el diario La Nación en Argentina, al momento en Ecuador la emplea Alcaldía de Quito.

5. Otras opciones adicionales

El análisis previo se centra en soluciones que están respaldadas por trayectoria y experiencia, cuentan además con presencia activa en el mundo de los datos abiertos. Sin embargo, la cantidad de plataformas disponibles resulta ser más amplio y siguen surgiendo otras

alternativas en el mercado, entre las que se referencian:

- DKAN: es un clon de CKAN basado en la plataforma de Drupal, confiriéndole flexibilidad siendo aprovechada por la comunidad existente detrás de la herramienta y su tecnología.
- Data.gov.uk To Go: es una herramienta proporcionada por el gobierno de Reino Unido puesto al servicio de usuarios en forma general.
- Data Press, proyecto en desarrollo del *Open Data Institute* cuya plataforma está basada en tecnología *open code* para la administración de datos “cloud” con base en Wordpress y CKAN.
- Open Data Soft, plataforma con servicios “en la nube” cuenta con cierto recorrido y alto número de usuarios, al momento se encuentra disponible en versión demo.

6. Comentarios finales

Considerando el surgimiento de los movimientos de acceso abierto en los ámbitos de la sociedad, lo que relaciona la gestión de la información, se encuentra inmerso en una época de cambios, donde intervienen normativas que permiten la reutilización de los datos para darle nuevos valores en lo social, cultural y económico. El contar con acceso a datos digitales con características *open data*, cuyos beneficios futuros dependen de su correcto uso, de igual forma dependerá de las estrategias, procesos y procedimientos orientados a la preservación digital que lleguen a adoptarse ahora, junto con el respaldo que puedan obtener de las autoridades o personeros de las organizaciones interesadas, que logren ser factores para su correcta aplicación y funcionalidad.

No obstante, se debe tener presente que los organismos públicos de investigación y la universidad, como tales, son quienes promueven el buen uso de estas plataformas para que lleguen a mayor público sin dificultades. Es imperiosa la necesidad de reconsiderar aspectos que presenten quienes serán los involucrados en el flujo de los datos; cuál será su campo de práctica; las necesidades y límites en cuanto a los datos que se desean compartir; qué tipos de datos se están generando, su forma de registro y calificación en relación a la apertura de los datos; la promoción y asesoría, de forma tal que estas representen oportunidades para fortalecer la sociedad del conocimiento.

7. Referencias Bibliográficas

Agenjo, X., & Hernández, F. (2010, October). Tendencias internacionales en el desarrollo funcional de la recuperación de la información: Linked Open Data (LOD). In *X Workshop Rebiun sobre proyectos digitales: diez años de proyectos digitales: cambian las bibliotecas, cambian los profesionales*. Valencia, 7 y 8 de octubre de 2010.

Aguilar, L. J. (2013). Computación en la nube: Notas para una estrategia española en cloud computing. *Revista del Instituto Español de Estudios Estratégicos*, 1(1).

Borrego, Àngel (2012). Los retos de la gestión de datos de investigación

Bucarey, S., & Álvarez, L. (2006). Metodología de construcción de objetos de aprendizaje para la enseñanza de anatomía humana en cursos integrados. *International Journal of Morphology*, 24(3), 357-362.

Cano-Parra, R. (2014) *Métricas y Reglas para la Escalabilidad Automática y Dinámica en la Nube Computacional Aplicadas en Entornos Educativos* (Doctoral dissertation, ThesisReport 004, Universidad de Valladolid, Valladolid, Spain

Cebrián, M (2009): “Comunicación interactiva en los cibermedios”. *Comunicar*, 33, 15-24. <http://dx.doi.org/10.3916/c33-2009-02-001>

Codina, L. (2003). La web semántica: una visión crítica. *El profesional de la información*, 12(2), 149-152

Ferrer-Sapena, A., Peset, F., & Aleixandre-Benavent, R. (2011). Acceso a los datos públicos y su reutilización: open data y open government. *El profesional de la información*, 20(3), 260-269.

García, F., Portillo, J., Romo, J., & Benito, M. (2007, September). Nativos digitales y modelos de aprendizaje. In *SPDECE*.

Lozano, G. V. (2008). Software libre.

Manovich, L(2013): *El software toma el mando*. Barcelona: UOCpres.

Martínez Méndez, F. J., & Rodríguez Muñoz, J. V. (2003). Síntesis y crítica de las evaluaciones de la efectividad de los motores de búsqueda en la Web. *Informationresearch*, 3(2), 8-2.

Ortega, J. L., & Aguillo, I. (2006). Análisis de co-enlaces: una aproximación teórica. *El profesional de la información*, 15(4), 270-277.

Peset, F., Ferrer-Sapena, A., & Subirats-Coll, I. (2011). Open data y linked open data: su impacto en el área de bibliotecas y documentación. *El profesional de la información*, 20(2), 165-174.

Ronda León, R. (2008). Arquitectura de Información: análisis histórico-conceptual. *No sólo usabilidad*, (7).

Sancho, J. V. (2014). Aproximación a una taxonomía de la visualización de datos.

Montuschi, L. (2001). Datos, información y conocimiento. De la sociedad de la información a la sociedad del conocimiento. *Serie Documentos de Trabajo de la Universidad del CEMA*, 192.

8. Otras Referencias

AENOR

https://www.aenor.es/aenor/normas/buscadornormas/buscadornormas.asp#.VTfgLGR_Oko

BOE, Norma Técnica de Interoperabilidad de Reutilización de recursos de información: <http://www.boe.es/boe/dias/2013/03/04/pdfs/BOE-A-2013-2380.pdf>

CKAN, the world's leading open-source data portal platform: <http://ckan.org/>

CKAN versión online de demostración <http://demo.ckan.org/es/>

Datos.gob.es, Guía para publicar datos abiertos de manera fácil y sencilla:

http://datos.gob.es/sites/default/files/guia-publicar-opendata-24h_v05.pdf

Directiva INSPIRE <http://www.ideo.es/europeo-inspire>

EC, DCAT Application Profile for data portals in Europe: https://joinup.ec.europa.eu/asset/dcat_application_profile/description

Junar, The Open Data platform: <http://www.junar.com/>

ODE (2011). Opportunities for data exchange. <http://www.alliancepermanentaccess.org/index.php/community/current-projects/ode>

OData (Open Data Protocol) https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=odata

Open DataSoft <http://public.opendatasoft.com/explore/?refine.language=en>

Socrata Open Data Portal: <http://www.socrata.com/products/open-data-portal/>

W3C, Data Catalog Vocabulary (DCAT): <http://www.w3.org/TR/vocab-dcat/>

Viavansi, OGoov - Plataforma de Gobierno Abierto, Datos Abiertos, Transparencia y Participación: <http://www.ogooov.com>

World Bank, Technical Assessment of Open Data Platforms for National Statistical Organisations: <http://documents.worldbank.org/curated/en/2014/10/20451797/technical-assessment-open-data-platforms-national-statistical-organisations>

ⁱVer sección "Compatibilidad con estándares" de este mismo documento.

ⁱⁱCKAN www.ckan.org

ⁱⁱⁱ Open Knowledge <https://okfn.org/>

^{iv} CMS http://en.wikipedia.org/wiki/Content_management_system

^vSu disponibilidad es proporcionada por medio de extensiones que servicios que son gratuitos lo utilizan por parte de terceros.

^{vi} Wordpress (es.wordpress.com)

^{vii}Componentes sobre los que se montan servicios de tele monitorización domiciliaria y componentes para ofrecer funcionalidades de trabajo cooperativo

^{viii} <http://marmotta.apache.org/>

^{ix} <http://skos.um.es/TR/rdf-sparql-query/>

^x <https://cloud.google.com/>

^{xi}W3C, Data Catalog Vocabulary (DCAT): <http://www.w3.org/TR/vocab-dcat/>

^{xii} <http://www.google.com/analytics/>

^{xiii}Una parte del portal de Chile está realizado con tecnología Junar.