

Plataforma de Datos Abiertos: una herramienta para promover la transparencia de la información gubernamental

Hugo Estrada, Javier Solís, Samuel Vieyra, Karen Nájera, Blanca Vázquez and Hasdai Pacheco
Fondo de Información y Documentación para la Industria INFOTEC
Av. San Fernando 37, Col. Toriello Guerra, Tlalpan, C.P. 14050, México D.F.
{hugo.estrada, javier.solis, samuel.vieyra, karen.najera, blanca.vazquez, ebenezer.sanchez}
@infotec.com.mx

Abstract—La iniciativa de Datos del Gobierno Abierto tiene el objetivo de promover el discurso civil, el mejoramiento del bienestar público y el uso eficiente de los recursos gubernamentales. Este objetivo se logra mediante la apertura de los datos del gobierno. Para llevar a cabo ésta apertura, es necesario contar con la infraestructura que permita gestionar y publicar los datos, para que éstos puedan ser accedidos por la ciudadanía. En este sentido, diversos países han apostado a la construcción de plataformas para exponer su información pública, estableciendo un canal de comunicación entre los gobiernos y sus ciudadanos. En este artículo presentamos una plataforma de Datos Abiertos desarrollada en México que integra un conjunto de funcionalidades para la gestión, publicación y consumo de datos públicos, con el objetivo de exponer la información gubernamental y proporcionar a los ciudadanos un espacio de consulta, participación y colaboración ciudadana.

Keywords—*Datos Abiertos, Plataforma de Datos Abiertos, transparencia.*

I. INTRODUCCIÓN

La iniciativa de Datos del Gobierno Abierto (OGD, por sus siglas en inglés)¹ es una iniciativa que promueve el discurso civil, el mejoramiento del bienestar público y el uso eficiente de los recursos públicos. De esta manera, busca ayudar a los ciudadanos a un mejor entendimiento de las actividades que realiza el gobierno y como éstas son realizadas, así como contribuir a la democracia de un país [1]. Los principales elementos de la OGD son: Datos Abiertos y Datos de Gobierno. Datos Abiertos se refiere a datos que pueden ser usados libremente, reusados y distribuidos por cualquier persona y Datos de Gobierno se refiere a cualquier dato producido por organismos públicos [1]. De esta manera, la iniciativa contempla la apertura de los datos gubernamentales. Con la finalidad de realizar correctamente la apertura de datos a los ciudadanos, dos organizaciones no gubernamentales, la Fundación Sunlight² y el grupo de trabajo en Gobierno Abierto³, definieron en el 2007, un conjunto de principios de datos del gobierno abierto para regir, guiar y dar soporte a la iniciativa de Datos Abiertos. Los principios declaran que los datos del gobierno deben ser considerados abiertos

únicamente si los datos se hacen públicos y cumplen con lo siguiente [2], [3]: los datos deben ser completos, primarios (sin procesamiento previo), oportunos, accesibles, procesables, no discriminatorios, presentarse en formatos no propietarios y ser publicados bajo una licencia abierta (disponibles sin restricciones de derechos de autor, mecanismos de control o patentes). Estos principios han sido aplicados por instituciones gubernamentales para publicar sus datos [4]. La tendencia de la iniciativa de Datos Abiertos está orientada a que los datos, además de que puedan ser consultados por los ciudadanos, puedan también ser legibles por los sistemas de software, y de esta manera, ser utilizados por aplicaciones desarrolladas por ciudadanos para explotar la información gubernamental. En este sentido, el creador de la Web Tim Berners-Lee propuso un esquema de cinco niveles de apertura de Datos Abiertos [5]. Actualmente este esquema es un estándar de la World Wide Web Consortium (W3C) [6]. Este organismo define que mientras más alto el nivel, mayor la apertura de los datos y mayores los beneficios para el ciudadano. La tabla I describe brevemente en que consiste cada uno de los niveles del esquema.

Actualmente, diversos países se han unido al movimiento de la apertura de su información pública así como la de sus procesos. Ejemplos de estos países son: Estados Unidos⁴, Reino Unido⁵, España⁶, Chile⁷, por mencionar algunos. Cada uno de estos países ha apostado a la construcción de plataformas de Datos Abiertos, cubriendo algunos niveles del esquema de apertura de Datos Abiertos. En este trabajo de investigación se presenta una plataforma de Datos Abiertos desarrollada por el Fondo de Información y Documentación para la Industria INFOTEC, México⁸ que tiene como objetivo exponer la información gubernamental y proporcionar a los ciudadanos un espacio de consulta, participación y colaboración ciudadana. Para ello, integra un conjunto de funcionalidades para la gestión, publicación y consumo de datos públicos. La plataforma se construyó con *SemanticWebBuilder* que es un framework de desarrollo dirigido por ontologías [7], por lo que una ontología

¹Open Government Data. <http://www.opengovdata.org/>

²Sunlight Foundation. <http://sunlightfoundation.com/>

³World Wide Web Consortium (W3C). <http://www.w3.org/Consortium/>

⁴<http://www.data.gov/>

⁵<http://data.gov.uk/>

⁶<http://datos.gob.es/>

⁷<http://datos.gob.cl/>

⁸INFOTEC. <http://www.infotec.com.mx/>

Tabla I. CINCO ESTRELLAS PARA LA APERTURA DE DATOS ABIERTOS LIGADOS [5]

Nivel de apertura	Descripción
1 estrella	Datos publicados en internet en formato no estructurado, por ejemplo, PDF.
2 estrellas	Datos publicados en internet en formato estructurado, por ejemplo: un archivo excel.
3 estrellas	Datos publicados en internet en formato estructurado y no propietario, por ejemplo: CSV o XML.
4 estrellas	Datos publicados en internet en formato estructurado no propietario, que además utilizan Identificadores de Recursos Uniformes (URI) y estándares abiertos de la W3C (RDF, SPARQL), por ejemplo: un dato que se accede a través de consultas de SPARQL.
5 estrellas	Datos publicados en internet en formato estructurado no propietario, que utilizan Identificadores de Recursos Uniformes (URI) y que además propician el ligado de los datos con otros datos para dotarlos de contexto.

es la base de la plataforma y funge como la estructura del sistema y el modelo de datos.

II. OBJETIVO

El objetivo de este trabajo de investigación es desarrollar una plataforma de Datos Abiertos soportada por tecnología semántica que permita: 1) transparentar las actividades del gobierno mediante la exposición de la información gubernamental representada como conjuntos de datos sin previo procesamiento (datasets); 2) proveer a la ciudadanía con medios para la consulta, descarga, utilización, valoración y socialización de los datasets; 3) proporcionar los mecanismos para facilitar el desarrollo de aplicaciones innovadoras que consuman y exploten los datos publicados en beneficio de la sociedad.

III. PLATAFORMA DE DATOS ABIERTOS

Como se mencionó anteriormente, la plataforma de Datos Abiertos desarrollada por INFOTEC está construida sobre *SemanticWebBuilder* que es un framework de desarrollo dirigida por ontologías [7]. Esto se refiere a que la base de la plataforma es una ontología y su objetivo es definir la estructura del sistema, el modelo de datos y el comportamiento de los objetos. Las principales ventajas de tener como base una ontología son i) puede ser fácilmente extendida para dar soporte a futuros requerimientos, ii) puede ser transformada automáticamente en código, de esta manera se genera el código de los objetos definidos y iii) se mantiene la persistencia de los datos al estar construida sobre una base semántica (definida en RDF). Una característica más de la plataforma es que actualmente ésta se encuentra en el nivel tres del esquema de apertura de Datos Abiertos [5]. Esto se refiere a que actualmente la carga de los datos se realiza tomando en cuenta formatos estructurados y no propietarios. Los formatos que se permiten cargar a la plataforma son: CSV, XML, KML, KMZ y JSON.

Para llevar a cabo la apertura de los datos gubernamentales, se definieron un conjunto de funcionalidades encargadas de la gestión y consumo de los conjuntos de datos públicos de las distintas instituciones, con el objetivo de proveer a los ciudadanos un espacio para la participación y colaboración. En las siguientes secciones, se describe cada una de las funcionalidades, así como los perfiles de usuario relacionados.

A. Perfiles de usuario de la plataforma de Datos Abiertos

Las funcionalidades de la plataforma están divididas de acuerdo al perfil de usuario. Los perfiles de usuario definidos son: a) publicador, b) consumidor (el cual se divide a su vez en consumidor registrado y consumidor no registrado) y c) administrador. A continuación se detalla cada uno de estos perfiles.

a) *Usuario publicador*: un publicador es una entidad (persona u organización) que tiene como objetivo principal la publicación de sus datasets en la plataforma para que otros usuarios tengan acceso a sus datos.

b) *Usuario consumidor*: un consumidor es toda entidad (persona o sistema informático) cuyo objetivo principal es explotar y hacer uso de los datos para informarse o generar cruces de información, reportes, aplicaciones, entre otros. Dentro de este perfil se consideran dos tipos:

- Consumidor no registrado o público: usuario no registrado en el sistema que puede visualizar y descargar cualquier dataset publicado en la plataforma, así como aplicaciones. Este usuario puede emitir comentarios y calificaciones tanto a datasets como a las aplicaciones que se encuentren disponibles.
- Consumidor registrado: usuario registrado en el sistema que además de contar con las mismas funcionalidades del consumidor no registrado, puede participar en el foro y colaborar con la plataforma a través de la carga de aplicaciones.

c) *Usuario Administrador*: se refiere al administrador de la plataforma. Un usuario administrador se encarga de los aspectos operativos del portal de datos abiertos y tiene acceso a todas las funcionalidades de la plataforma.

B. Funcionalidades de la plataforma de Datos Abiertos

Las funcionalidades de la plataforma están categorizadas en cuatro módulos, los cuales son: a) Gestión de datasets, b) Gestión de aplicaciones, c) Gestión de interacción social y d) Gestión de la plataforma de Datos Abiertos. A continuación se describe cada uno de los módulos así como las funcionalidades más relevantes de cada uno de estos.

a) *Gestión de datasets*: es el módulo encargado de las funcionalidades relacionadas con la gestión y control de los datasets. Los perfiles de usuario que tienen acceso a este módulo son: publicador, consumidor y administrador. Las funcionalidades relevantes de la gestión de datasets se describen a continuación.

- Carga de datasets: gestiona la carga de datasets en formato no propietarios (CSV, XML, KML, KMZ, JSON). Cada dataset debe ser descrito a través de un conjunto de metadatos previamente definidos, por ejemplo: nombre, descripción, etiquetas relacionadas al dataset.
- Búsqueda de datasets: permite la búsqueda de datasets específicos. La búsqueda de datasets se realiza por palabra clave usando los metadatos: ID, título, etiquetas y descripción.

- Vista a detalle de un dataset: permite la visualización de los detalles de un dataset. En esta vista se muestran todos los metadatos que describen al dataset y que fueron capturados durante la carga del dataset.

b) Gestión de aplicaciones: es el módulo encargado de las funcionalidades relacionadas con la gestión y control de aplicaciones que consumen los datos alojados en la plataforma. Los perfiles de usuario que tienen acceso a este módulo son: consumidor y consumidor no registrado. Las funcionalidades relevantes de la gestión de aplicaciones se describen a continuación.

- Carga de aplicaciones: gestiona la carga de una aplicación (y sus artefactos) que explote el contenido de por lo menos un dataset.
- Búsqueda de aplicaciones: permite la búsqueda de aplicaciones específicas. La búsqueda de datasets se realiza por palabra clave usando los metadatos: título, categorías y descripción.
- Vista a detalle de una aplicación: permite la visualización de los detalles de una aplicación. En esta vista se muestran todos los metadatos que describen a la aplicación y que fueron capturados durante la carga.

c) Gestión de interacción social: es el módulo encargado de las funcionalidades relacionadas con la gestión y control del foro y comentarios realizados por los usuarios en la plataforma. A su vez, gestiona la interacción de la plataforma con redes sociales como Twitter, Facebook y Google+. Todos los perfiles de usuario tienen acceso a este módulo. Las funcionalidades relevantes de la gestión de interacción social se describen a continuación.

- Comentar datasets y aplicaciones: permite a un usuario consumidor realizar comentarios en datasets y aplicaciones. Los comentarios se pueden realizar desde la vista a detalle del dataset o de la aplicación.
- Calificar datasets y aplicaciones: permite otorgar una calificación tanto a datasets como a aplicaciones. Esta puntuación es a través de una escala de 1 a 5 estrellas (1=No recomendable; 2=Regular; 3=Bueno; 4=Recomendable; 5=Excelente).
- Participación en el foro: permite al consumidor participar en el foro del portal de Datos Abiertos que propicia la interacción con los usuarios consumidores. El foro está dedicado para que los consumidores planteen sugerencias, dudas, comentarios, quejas del portal, entre otros.

d) Gestión de la plataforma de Datos Abiertos: este módulo se encarga de las funcionalidades relacionadas con la gestión de los contenidos, el registro, autenticación y administración de usuarios, además de validar el inicio de sesión de los usuarios a través del servicio de firma electrónica. Los perfiles de usuario que tienen acceso a este módulo son: consumidor, administrador y publicador. Las funcionalidades relevantes de la gestión de la plataforma de Datos Abiertos se describen a continuación.

- Registro de una institución: permite el registro de una institución (publicador). Cada institución debe describirse a través de un conjunto de metadatos previamente definidos, por ejemplo: nombre, descripción, siglas.
- Administración de usuarios con perfil publicador y consumidor: permite el registro, eliminación y autenticación de publicadores y consumidores, así como administrar la información específica de cada uno de ellos.
- Administración de los contenidos del portal: consiste en la creación y edición de contenidos estáticos del portal de Datos Abiertos.

En la tabla II se presenta una lista de todas las funcionalidades que cubre la plataforma de Datos Abiertos, las cuales están clasificadas de acuerdo al módulo de gestión al que pertenece y al usuario que tiene acceso a este módulo.

Como parte de este trabajo de investigación, se llevó a cabo un análisis comparativo de las funcionalidades de portales de datos abiertos. El objetivo de este análisis fue determinar si nuestra plataforma se alinea con las buenas prácticas detectadas por otras plataformas, así como demostrar las funcionalidades ausentes o adicionales que ésta cuenta con respecto a otras, finalmente con este análisis también se identifica en que casos es necesario el registro del consumidor. Nuestra plataforma de datos abiertos fue comparada con las plataformas desarrolladas por los países de: Estados Unidos (Data.gov), Reino Unido (Data.gov.uk), España (Datos.gob.es) y Chile (Datos.gob.cl).

En la tabla III se presentan los resultados de este análisis. En la primera columna se presentan las funcionalidades detectadas en los portales de datos abiertos. Las siguientes columnas indican por país, si se encontró la funcionalidad o no en el portal y el tipo de usuario consumidor que tiene acceso a la funcionalidad: registrado, aquel consumidor registrado en el portal y no registrado, aquel consumidor no registrado en el portal. La tabla III tiene algunas anotaciones las cuáles se describen a continuación: se indica con un asterisco (*) aquella funcionalidad en donde se requiere ingresar al portal el nombre y/o email de la persona que realiza la acción, el texto “No aplica” se refiere a una funcionalidad no existente en el portal y con el uso de guión (-) se refiere a que no se encontró información de la funcionalidad.

IV. CONCLUSIONES Y TRABAJO FUTURO

En el presente artículo se describió una plataforma de Datos Abiertos desarrollada en INFOTEC que tiene como objetivo permitir la exposición de información gubernamental y proveer a la ciudadanía con medios para la consulta, descarga, utilización, valoración y socialización de los datasets; así como con mecanismos para facilitar el desarrollo de aplicaciones innovadoras que consuman y exploten los datos publicados en beneficio de la sociedad. La plataforma está conformada por cuatro módulos de gestión: Gestión de datasets, Gestión de aplicaciones, Gestión de interacción y Gestión de la plataforma, de los cuales se describieron sus funcionalidades más relevantes. De igual manera, se describieron los perfiles

Tabla II. FUNCIONALIDADES DE LA PLATAFORMA DE DATOS ABIERTOS

Módulo de gestión	Funcionalidades	Perfil de usuario
Gestión de datasets	Carga, edición, desactivación, eliminación, descarga, listado, búsqueda, trazabilidad y vista a detalle de un dataset.	Publicador, consumidor, administrador.
Gestión de aplicaciones	Carga, búsqueda, vista a detalle, descarga, galería y edición de una aplicación.	Consumidor, consumidor registrado.
Gestión de interacción social	Comentar, calificar, monitoreo de comentarios y compartir datasets y aplicaciones en redes sociales. Participación en el foro, administración del foro.	Consumidor, consumidor registrado, administrador
Gestión de la plataforma de Datos Abiertos	Registro de usuarios con los perfiles administrador, consumidor y publicador. Inicio de sesión y autenticación del administrador, del publicador y consumidor. Registro de una institución, administración (altas, bajas y modificaciones) de catálogos: temas e instituciones. Administración de usuarios con perfil publicador y consumidor. Administración de los contenidos del portal. Estadísticas de uso de datasets y aplicaciones.	Administrador, consumidor, publicador

Tabla III. ANÁLISIS COMPARATIVO DE FUNCIONALIDADES DE LA PLATAFORMA DE DATOS ABIERTOS

Funcionalidad	Estados Unidos	Reino Unido	España	Chile	Nuestra plataforma
Descarga de datasets	No registrado	No registrado	No registrado	-	No registrado
Búsqueda de datasets	No registrado	No registrado	No registrado	No registrado	Registrado
Vista a detalle de un dataset	No registrado	No registrado	No registrado	No registrado	No registrado
Cargar una aplicación	Registrado	No registrado*	Registrado	No aplica	Registrado
Búsqueda de aplicaciones	No registrado	No registrado	No registrado	No registrado	Registrado
Descarga de aplicaciones	-	-	-	-	No registrado
Comentar un dataset	Registrado	Registrado	No registrado*	-	No registrado*
Comentar una aplicación	Registrado	Registrado	No registrado*	-	No registrado*
Calificar un dataset	No aplica	No aplica	No registrado	No aplica	No registrado
Calificar una aplicación	No aplica	No aplica	No registrado	No aplica	No registrado
Comentar en un foro	Registrado	Registrado	No aplica	-	No registrado

de usuario soportados por la plataforma y las funcionalidades a las que cada perfil tiene acceso. También se presentó un análisis comparativo que muestra las diferencias y similitudes de nuestra plataforma en desarrollo actual con respecto a otras plataformas. Las funcionalidades analizadas fueron obtenidas a partir de examinar exhaustivamente cada una de las plataformas descritas. De esta manera, se muestra que las funcionalidades de nuestra plataforma en su versión actual están alineadas con las funcionalidades desarrolladas por otros países y además todas son cubiertas.

Actualmente la plataforma se encuentra en el nivel tres del esquema de apertura de Datos Abiertos [5]. Como trabajo futuro se extenderá la plataforma para alcanzar los niveles de apertura 4 y 5, lo cual permitirá lo siguiente: i) representar los datos en formatos estándares de la W3C como OWL y RDF, habilitando el uso de razonadores para inferir infor-

mación; ii) vincular los datos expuestos en la plataforma con datasets externos para ampliar el contexto de la información, por ejemplo: los nombres de municipios pueden enlazarse con las definiciones de los mismos presentadas en DBpedia⁹ o Geonames¹⁰; iii) posicionar a nuestra plataforma a nivel internacional como una plataforma de Datos Abiertos que permite publicar datos gubernamentales, generar información automáticamente a partir de ellos y a su vez dotarlos de contexto.

REFERENCES

- [1] U. B, "Open government data: Towards empirical analysis of open government data initiatives," *OECD Working Papers on Public Governance*, no. 22, pp. 127–139, 2013.
- [2] Sunlight Foundation, "Guidelines for Open Data Policies. Online," <http://sunlightfoundation.com/opendataguidelines/>, 2013, last access: 05/08/2013.
- [3] W3C, "Straightforward Steps to Publish Government Data. Online," <http://www.w3.org/TR/gov-data/#steps>, 2009, last access: 05/08/2013.
- [4] D. Lathrop and L. Ruma, *Open Government: Collaboration, Transparency, and Participation in Practice*, 2010.
- [5] T. Berners-Lee, "Linked Data," <http://www.w3.org/DesignIssues/LinkedData.html>, 2009, last access: 23/08/2013.
- [6] D. Bennett and A. Harvey, "W3C Open Government Data," <http://www.w3.org/TR/gov-data/>, 2013, last access: 05/07/2013.
- [7] J. Solis, H. Pacheco, K. Najera, and H. Estrada, "A MDE Framework for Semi-automatic Development of Web Applications," in *Proceedings of the 1st International Conference on MDE and Software Development*, 2013, pp. 241–246.

⁹DBpedia. <http://www.dbpedia.org/>

¹⁰Geonames. <http://www.geonames.org/>