

## **Empoderamiento de los ciudadanos en el análisis de datos abiertos**

Raúl Martínez

Director: David Bañeres Besora

### **Abstract**

La disciplina del análisis de datos ha formado parte de la sociedad durante las últimas décadas. Su aplicación ha trascendido a múltiples disciplinas, incluyendo medicina o informática, ayudando a miles de profesionales a tomar decisiones correctas en base a unas hipótesis apoyadas por una recogida de datos. A pesar de ir cobrando cada vez más importancia, hasta el punto de que la administración pública ha empezado a poner a disposición de la sociedad dichos datos, no existe una forma clara de explotar dichos datos para personas profanas en la disciplina.

Este artículo pretende dar, por un lado, una visión global del estado del arte en lo que a análisis de datos abiertos se refiere y, por otra, de la forma de acceso a dichos datos por parte del sector de la sociedad que no posee conocimientos informáticos avanzados. Es por este motivo que se realiza un caso de uso para ayudar a las personas sin conocimientos informáticos a poder acceder a datos abiertos y a su posterior análisis.

### **Palabras clave**

Visualización de datos, análisis de datos, oData, citizen empowerment.

## 1. Introducción

El término *análisis de datos* fue acuñado por John Tukey por vez primera en 1961 en su obra 'The future of data analysis' (Tukey, 1962) como "Procedimientos para analizar datos, técnicas para interpretar los resultados de dichos procedimientos, formas de planear la recolecta de datos para hacer el análisis más fácil, más preciso o más exacto".

El objetivo del análisis de datos es obtener información útil, que aporte valor, de forma que se obtengan conclusiones que apoyen una posterior toma de decisiones. Esta disciplina surgió como una necesidad para entender el porqué de las cosas y como poder solucionarlo mediante la interpretación de los datos.

El análisis de datos se ha realizado tradicionalmente por personal con conocimientos informáticos, lo que ha dificultado el acceso a la sociedad en general. Es por esto que el auge de los datos en abierto en la actualidad cobra especial importancia, y por ello es necesario un proceso de acercamiento de dichos datos hacia el público en general.

En este artículo se procederá a realizar un apartado en el que se introducirá al lector en la disciplina de los datos abiertos, haciendo un repaso de su historia, sus características (apartado 2). Se analizarán los beneficios y las barreras a la hora de utilizar los datos abiertos en la sociedad (apartado 3) y cómo esto contribuye al empoderamiento de los ciudadanos, además del proceso para conseguir analizar datos (apartado 4). Posteriormente se explicará un caso de uso para la herramienta creada (apartado 5) y finalmente se expondrán las conclusiones derivadas de la investigación realizada (apartado 6).

## 2. Datos abiertos

Los datos abiertos, u *open data*, es una práctica que persigue que ciertos datos estén disponibles de forma libre para todo el mundo, sin restricciones de derechos de autor, patentes u otros mecanismos de control. Se basa en tres principios fundamentales: amplitud de miras (*openness*), participación y colaboración. (Huijboom, 2011).

Aunque el término fue acuñado por primera vez en un informe de 1995 de una agencia científica americana<sup>1</sup>, no fue hasta el 2007 cuando unos científicos se reunieron en Sebastopol (Estados Unidos) para discutir el concepto de datos públicos abiertos. En esa reunión se acordó la utilidad que proponía el uso de esos datos para la sociedad y es dónde se empezó a utilizar el término de *open data* tal y como lo conocemos en la actualidad.

Existen dos tipologías principales de datos, según sean privados o públicos, en función de si el dueño es un individuo (privado) o es un estamento público (público).

Para que unos datos sean considerados como *abiertos* deben ser: públicos, detallados, actualizados, accesibles, automatizados, sin registro, abiertos y libres (Bauer & Kaltenböck, 2011).

Todos los datos de carácter público que no entrañen restricciones de privacidad, seguridad o derechos de autor deben ser publicados. Deben ser en bruto, esto es, tal y como están en su origen, sin ser procesados de ninguna manera y deben ser suministrados a los usuarios con la frecuencia necesaria para que sean precisos y actuales. Además deben poder ser consultados por el mayor número de usuarios posible. Por lo tanto, no debería existir ninguna restricción para que cualquier persona pueda hacer uso de dichos datos, sea cual sea su objetivo de uso. Deben cumplir una estructura de forma que puedan ser procesados automáticamente por un ordenador. Esta condición contribuirá a que se puedan reutilizar los datos de una forma automática. Los datos además de estar disponibles para todos, sin necesidad de identificarse previamente, deben ser no propietarios, es decir, no pueden depender de una entidad o de una herramienta propietaria de una entidad. Así, pertenecen a datos no propietarios los formatos XML o JSON, mientras que en el lado de los formatos propietarios se ubican, por ejemplo, los pertenecientes a la *suite* Office de Microsoft: XLS (Excel) o DOCX (Word). Por último, los datos deben ser de uso 100% libre para los usuarios, libres de derechos, patentes, copyright y no estar sujetos a derechos de

---

1

<http://parisinnovationreview.com/articles-en/a-brief-history-of-open-data>

privacidad, seguridad o privilegios que puedan estar condicionados por otras normas.

Conviene destacar la clasificación realizada por Tim Berners-Lee en 2006 (Berners-Lee, 2006), la cual clasifica desde una a cinco estrellas según sea la calidad de los datos. En el caso de tener una estrella, los datos se ofrecen en soporte papel, siendo muy difícil su explotación. Si se encuentran en un documento pero se necesita una herramienta propietaria para utilizarlo, hablamos de dos estrellas. Nuestros datos contarán con tres estrellas cuando se hallen dentro de un documento y además se puedan utilizar herramientas no propietarias para su edición. Cuando los datos se encuentren correctamente estructurados y accesibles por URL hablaremos de cuatro estrellas y, cuando además de lo anterior, estén enlazados entre sí, podremos obtener la máxima clasificación de cinco estrellas. En este último punto de la clasificación se hace énfasis en un concepto, el llamado *linked data*.

### 3. Empoderamiento de la sociedad

Los beneficios derivados de la utilización de los datos abiertos se producen a tres niveles principalmente. (Janssen et al., 2012): políticos y sociales, económicos y operativos.

Los beneficios políticos y sociales incluyen transparencia, empoderamiento de los ciudadanos, mejora de los procesos, además de las mejoras de los servicios ofertados a los ciudadanos por parte del gobierno. Socialmente se produce una estimulación del conocimiento, nuevos servicios sociales (y mejora de los ya existentes) y creación de nuevas oportunidades en el sector.

En cuanto a los beneficios económicos encontramos la estimulación del conocimiento y la innovación, el desarrollo de nuevos productos y servicios y el crecimiento económico derivado de la estimulación de la competencia. Especialmente importante también es la disponibilidad de información para inversores y empresas.

Finalmente, a otro nivel se encuentran los beneficios operativos, incluyendo la reutilización de los datos, la optimización de procesos administrativos, la obtención de nuevos datos a partir de la combinación de otros y la sostenibilidad de los mismos. También se vuelve especialmente interesante

el hecho de poder integrar datos públicos y privados.

A su vez, existen cinco mitos que conviene desmontar acerca del open data (Janssen et al., 2012). Que la publicación de los datos en abierto retornará beneficios instantáneos, como es obvio, dependerá del uso que se haga de los mismos, puesto que si el análisis no retorna una información válida, no permitirá que se generen los beneficios deseados. Por otro lado, la información no debería publicarse de manera indiscriminada, sino que debería pasar un filtro para dilucidar qué es lo que merece ser o no liberado. Del mismo modo, y enlazando con el siguiente mito, siempre hay que verificar la calidad de los datos y su adecuación, de forma que no sirva cualquier dato para ser publicado, sino que es posible que sea necesaria una adecuación de los mismos para poderlos hacer explotables y analizables correctamente. Otro de los mitos es que toda persona pueda ser capaz de utilizar dichos datos, ya que, nuevamente, dependerá de diversos factores como educación o sector en el que vaya a analizar dichos datos. Por último, y no menos importante, hay que desmitificar el hecho de que cualquier dato abierto va a desembocar en un gobierno abierto. Implantar un gobierno abierto implicaría una transformación considerable en el sector público, transformación que, por sí sola, los datos abiertos no pueden crear, sino que necesitan implicaciones a otros niveles, políticos sobretodo.

Después de observar los beneficios y los mitos comentados, dependerá de cada caso concreto la implantación o no de los datos abiertos para la sociedad.

La publicación de datos abiertos es una tendencia al alza, si tenemos en cuenta el informe creado por Open Data Barometer<sup>2</sup>. Existen diversos estamentos que promueven el uso de datos abiertos, como por ejemplo el Open Data Institute<sup>3</sup>, entidad sin ánimo de lucro que lucha por la difusión y uso de los datos abiertos.

Cada vez más, las instituciones de la administración se deciden a liberar datos en

---

2

[http://opendatabarometer.org/?\\_year=2016&indicator=OD](http://opendatabarometer.org/?_year=2016&indicator=OD)

<sup>3</sup> <https://theodi.org>

forma de datos abiertos, contribuyendo al empoderamiento del conjunto de la ciudadanía (también conocido como *empowerment of citizens*), lo que nos hace poder empezar a hablar de otros conceptos como *open government*, tal como comentan (Ferrer-Sapena et al., 2011). Así, las administraciones centrales se deciden a incorporar aplicaciones tanto móviles (Sandoval et al., 2012) como de escritorio las cuales permiten que los ciudadanos se sientan más cerca del gobierno y obtengan un mayor conocimiento tanto de este como de la sociedad que habitan. En este sentido también se desarrollan herramientas (Villena-Román, et al. 2014) que buscan determinados patrones de comportamiento de los ciudadanos en las redes sociales como Twitter así como diferentes portales de datos abiertos.

El último aspecto a tener en cuenta es el hecho de que esta liberación de datos permite que los ciudadanos se puedan convertir en desarrolladores y por tanto en analistas. Existen APIs disponibles para que cualquier persona pueda recoger dichos datos y realizar sus propios análisis, como Aemet <sup>4</sup> o incluso al Portal de datos abiertos de la UE <sup>5</sup>.

La influencia del análisis de datos se ha propagado a través de diferentes disciplinas, tales como la educación, donde cobra especial importancia, permitiendo realizar análisis de datos y análisis predictivo de los mismos, es lo que se conoce como *learning analytics*. Uno de los principales factores que hacen que el *learning analytics* se haya transformado en los últimos años es el hecho de disponer de datos educativos preparados para ser analizados. Otros factores que han contribuido han sido la mayor capacidad computacional para poder explotar dichos datos, el desarrollo de comunidades educativas o la creación de herramientas específicas para la explotación de dichos datos.

En relación a otras disciplinas como la física o la biología, las ciencias del aprendizaje son relativamente nuevas en cuanto a la analítica de datos. No fue hasta 1970 cuando una publicación (*Computers in Biology and Medicine*) empezó a escribir sobre la analítica aplicada a la biología (Baker, Siemens 2014).

<sup>4</sup> <https://opendata.aemet.es/>

<sup>5</sup> <http://data.europa.eu/euodp/es/developerscorner>

Especial importancia cobra en la disciplina de la medicina, dónde se encuentran estudios como por ejemplo en genética (Weir, 1990, Pope et al., 2000) o en áreas más específicas como cuidado de la salud (Bradley et al., 2007).

En la actualidad, multitud de instituciones educativas se ocupan cada vez más de intentar predecir mediante modelos, el comportamiento de sus estudiantes. Universidades como la UOC<sup>6</sup> o la UPC<sup>7</sup> por nombrar algunas, ya dedican parte de su esfuerzo a intentar entender a qué se debe determinado comportamiento del alumnado en cuanto a técnicas de estudio o tasa de abandono.

En los últimos años, con el auge del big data están tomando especial importancia las técnicas de inteligencia de negocio (en inglés *business intelligence*). El *big data* consiste básicamente en la obtención de grandes cantidades de datos, los cuales, analizados de la manera correcta, se transforman en un subconjunto de dichos datos que permiten a la entidad que los posee determinar patrones para posteriormente, tomar decisiones empresariales.

Existe una relación directa entre *business intelligence* y *big data (open data)*, por lo que a la fuerza, obtener mayor cantidad de datos proporciona una inteligencia de negocio mayor (Chen, et al, 2012). En la actualidad, el uso de big data está siendo aplicado por ejemplo a la medicina y está consiguiendo, entre otras cosas<sup>8</sup> crear análisis predictivo o planificación estratégica, entre otras. Es esta relación entre *big data* y *open data* lo que puede permitir que se puedan producir importantes avances en un área tan importante como es la medicina, pero también en otras.

Además de las ya comentadas influencias en las disciplinas anteriores cabe hacer hincapié en la posibilidad que ofrece el análisis de datos: los llamados métodos predictivos. Dichos método tienen como propósito prever o

<sup>6</sup> <http://www.uoc.edu/portal/es/elearncenter/learning-analytics/index.html>

<sup>7</sup> <https://inlab.fib.upc.edu/es/learning-analytics>

<sup>8</sup> <https://www.datapine.com/blog/big-data-examples-in-healthcare/>

anticipar situaciones futuras, y para ello requiere de la exploración, la descripción, la comparación, el análisis y la explicación. El propósito principal de una investigación de este tipo es “predecir” la dirección futura de los eventos investigados. Así, se encuentran modelos predictivos de todo tipo, desde redes neuronales aplicadas a la meteorología y la oceanografía (Hsieh & Tang, 1998) o la bancarrota en el sector de la banca (Odom & Sharda, 1990).

#### 4. Proceso de análisis de datos

El proceso de análisis de datos se divide en tres partes principales: captura, procesamiento y consulta.

La captura de los datos comprende la recogida de los datos, que como veremos en la última parte (la correspondiente a la consulta), pueden proceder de diversas fuentes. Para obtener dichos datos se pueden utilizar técnicas como el *web scraping*, o rascado de datos. Esta técnica simula la navegación como si fuera un humano mediante un programa informático, ya sea utilizando el protocolo HTTP o manualmente, de forma que vaya recorriendo la página web para guardar todos los datos que en ella se encuentren para guardarlos en un almacenamiento de datos, de dónde se procesarán en la siguiente fase del análisis. Otra forma de obtener los datos puede ser mediante peticiones a APIs de las entidades que exponen los datos o por simple descarga desde las propias páginas web.

El procesamiento consiste en acumular y manipular elementos de los datos capturados para poder producir información analizable. Los datos se pueden procesar manualmente, automáticamente o electrónicamente. El procesamiento de datos electrónico se refiere por ejemplo a tarjetas perforadas utilizadas para el censo de Estados Unidos de 1890. Por último, el procesamiento automático es el más utilizado en la actualidad, en el que un ordenador (o servidor, según el caso), es el encargado de recoger y transformar dichos datos.

Para poder analizar datos, estos necesariamente debe tener un formato concreto, ya sea para visualizarlos o bien para publicarlos.

Aún existiendo multitud de formatos, hay tres que destacan por su gran volumen de uso:

- CSV formatea los datos separados por comas, de forma que puedan ser leídos por la mayoría de herramientas actuales.
- *XML (Lenguaje Etiquetado Extensible)* también se considera un estándar en cuanto a comunicación entre plataformas. De este lenguaje deriva *El RDF (infraestructura de definición de recursos)*, el cuál facilita el intercambio de recursos entre plataformas.
- *oData (Opendata protocol)*, permite generar consultas y subsecciones entre los conjuntos de datos para después, poderlos exportar. Es un protocolo en auge gracias al respaldo de gigantes de la industria informática como Microsoft.

La disciplina de la analítica de datos comprende la visualización y la publicación de los mismos. Existen diversas formas de visualizar datos en función de la forma en la que se ofrezcan dichos datos.

Existen determinados estamentos que exponen datos directamente en su web, por lo que el usuario, simplemente se conecta a dicha web y realiza la visualización (o incluso la descarga de los conjuntos de los datos, si procede) desde allí.

También se pueden analizar mediante conjunto de rutinas que permiten la interconexión entre varias piezas de software (también conocidas como APIs). Los datos no son accesibles directamente desde la web sino que el usuario debe realizar una petición a la API, la cual se los devuelve en un formato previamente pactado para que el usuario pueda hacer uso de los mismos.

La diferencia principal entre esta forma de consumir datos y la anterior es el entorno en el que se consumen, ya que aquí el usuario necesita poseer de una herramienta externa a la entidad expositora de los datos para poder realizar el análisis.

Otra opción es mediante *datasets*, también denominados conjuntos de datos. Esta opción presupone que los datos ya están liberados en un fichero que el usuario puede descargar directamente a su ordenador y empezar a analizar, ya sea con alguna herramienta

propietaria, no propietaria o con una aplicación de análisis desarrollada a medida por terceros.

Por último, se ofrece la opción de analizar los datos mediante herramientas específicas, como por ejemplo *D3js*<sup>9</sup> (librería de Javascript que permite, de forma programática, explorar conjuntos de datos y mostrarlos de una forma visualmente atractiva. Así, el usuario puede realizar gráficos de diferentes tipos, incluyendo de barras, de sector, de dispersión, etc.); *Google public data explorer*<sup>10</sup> (la alternativa de Google a los visualizadores de datos convencionales) o *Eclipse BIRT*<sup>11</sup> (permite crear informes para posteriormente, incluirlos en páginas web).

Respecto a la publicación de datos abiertos, esto es convertirse en la entidad que libera los datos y no la que los consume, existen diversas opciones en forma de plataforma que pueden ayudar a distribuir dichos datos en la red. Estas plataformas son *OpenDataSoft*<sup>12</sup> (comercial) y *cKan.org*<sup>13</sup> (*open source*) permiten subir *datasets* a internet y definir una API que los exponga en pocos minutos. De esta forma se consigue que los datos, además de públicos, sean accesibles para todo el mundo.

Existen diversas tipologías de datos en relación a la formación de los mismos. Por un lado, encontramos datos estructurados, los que se corresponden a una estructura clara y entendible, y que por tanto, estén disponibles para su uso sin necesidad de realizar un limpiado de los mismos. Por el otro, existen los denominados datos desestructurados: estos son los que requieren de un proceso de limpiado y estructurado para poder ser utilizados correctamente en un proceso de análisis.

El hecho de poder acceder a fuentes de información proporcionadas por las administraciones públicas, puede contribuir al conocido fenómeno de *empowerment of citizens*. Dicho fenómeno habla de la capacidad de un miembro, en este caso un ciudadano, para contribuir al desarrollo de una comunidad y de hacerla más próspera.

<sup>9</sup> <https://d3js.org/>

<sup>10</sup> <https://www.google.com/publicdata/>

<sup>11</sup> <http://www.eclipse.org/birt/>

<sup>12</sup> <https://www.opendatasoft.es/>

<sup>13</sup> <https://ckan.org/>

Llegados a este punto cabe discernir entre el análisis de datos llevado a cabo por profesionales de la informática y el de otro tipo de profesiones que no posean una base en programación informática. En este sentido, se hace especialmente importante el hecho de que un ciudadano, sin habilidades para el análisis de datos sea capaz de utilizarlos y de poder realizar sus propios análisis.

La investigación de las herramientas existentes se ha subdividido según sea adecuado para perfiles programadores y/o no programadores. A su vez, se ha discriminado entre soluciones de pago o por el contrario, *open-source*<sup>14</sup>. Existe multitud de oferta en relación a herramientas y dependiendo del objetivo de la investigación o del sector de los datos puede ser más o menos conveniente una herramienta u otra.

En el apartado de herramientas para programadores encontramos en el apartado de comerciales como principal a *Matlab*<sup>15</sup>, el cual se presenta como una solución adecuada tanto para empresa como para ámbito académico. En la versión no comercial encontramos *JMatlab*<sup>16</sup>, muy parecido a su homónimo de pago y otras suites basadas en los lenguajes de programación Python y R, como son *DataMelt*<sup>17</sup> y *GNUOctave*<sup>18</sup>, respectivamente.

En cuanto a las opciones disponibles para no programadores encontramos *Minitab*<sup>19</sup> o *SPSS*<sup>20</sup> como alternativas comerciales y *Tableau*<sup>21</sup>, la cual dispone de un seguimiento muy importante según demuestran los datos de Google Trends<sup>22</sup>, gracias a su menor curva de aprendizaje en relación a sus competidores. En la vertiente de las herramientas no comerciales encontramos

<sup>14</sup>

<https://conceptosclaros.com/19-herramientas-de-analisis-de-datos>

<sup>15</sup> <https://es.mathworks.com/products/matlab.html>

<sup>16</sup> <http://jwork.org/jmathlab/>

<sup>17</sup> <http://jwork.org/dmelt/>

<sup>18</sup> <https://www.gnu.org/software/octave>

<sup>19</sup> [www.minitab.com](http://www.minitab.com)

<sup>20</sup> <https://www.ibm.com/es-es/marketplace/spss-statistics>

<sup>21</sup> [www.tableau.com](http://www.tableau.com)

<sup>22</sup>

<https://trends.google.es/trends/explore?q=tableau.spss.minitab>



como principales representantes a *Rcommander*<sup>23</sup> y *Rapid Minner*<sup>24</sup>.

El análisis de datos y su explotación se vuelve especialmente complicado para personas que no poseen una base informática a nivel de programación. Ya sea a la hora de obtener dichos datos o a la hora de incorporarlos a una herramienta, no existen demasiadas opciones que faciliten el trabajo para personas no iniciadas.

Pese a existir herramientas comerciales (*SPSS* o *SaaS*) y otras desarrolladas por investigadores (De Cáceres et al., 2003) en cuyas últimas versiones se vuelven más fáciles de usar (incluso para personas sin demasiados conocimientos dentro del ámbito), parece claro que continúa siendo necesaria una base técnica consolidada para poder visualizar y analizar datos, por lo que se hace necesario seguir investigando en este sentido. Se constata que las herramientas existentes presuponen que el usuario es capaz de entender cómo funciona la herramienta y además, posee conocimientos de estadística suficientes para poder explotar dichos datos.

## 5. Caso de uso: WordPress para citizens

*oData*<sup>25</sup> es un protocolo de datos abiertos (también conocido como Open Data Protocol) que define un conjunto de buenas prácticas para construir y consumir API RESTful. Surgió en 2007 gracias Microsoft y permite la creación y el consumo de dichas APIs, las cuales que permiten a los clientes web publicar y editar recursos, usar URLs y definirse en un modelo de datos, usando mensajes HTTP simples.

El esquema típico de una petición *oData* es: <URL servidor>/Raíz del servicio/Ruta del recurso/?Parámetros de consulta.

En la siguiente URL, la raíz del servicio es *OData/OData.svc/*, la ruta del recurso es *Category(1)/Products* y los parámetros de consulta son *\$top=2&\$orderby=name*

[http://services.odata.org/OData/OData.svc/Category\(1\)/Products?\\$top=2&\\$orderby=name](http://services.odata.org/OData/OData.svc/Category(1)/Products?$top=2&$orderby=name).

Para este caso concreto, estaríamos diciéndole al sistema que devolviese los productos pertenecientes a la categoría 1 y

además, los dos primeros ordenados por nombre.

El protocolo de *oData* se compone de un conjunto de interacciones RESTful, las conocidas como CRUD (Create, Read, Update y Delete - crear, leer, actualizar y borrar), junto con un lenguaje definido de consultas a datos. Los datos enviados por el servicio *oData* pueden ser representados en XML basado en el protocolo Atom/AtomPub<sup>26</sup> o en JSON.

Para personal no informático resulta complejo trabajar directamente con llamadas REST y protocolo *oData*. Por este motivo se propone desarrollar una herramienta que permite obtener dichos datos sin necesidad de conocimientos informáticos avanzados.

El navegador interactúa con la aplicación que va a realizar la petición (en este caso WordPress), la cual mediante el uso del protocolo *oData* consulta un servicio *oData*, que a su vez se encarga de realizar la consulta a la base de datos, devolviendo los datos solicitados nuevamente al navegador. (Fig. 1)

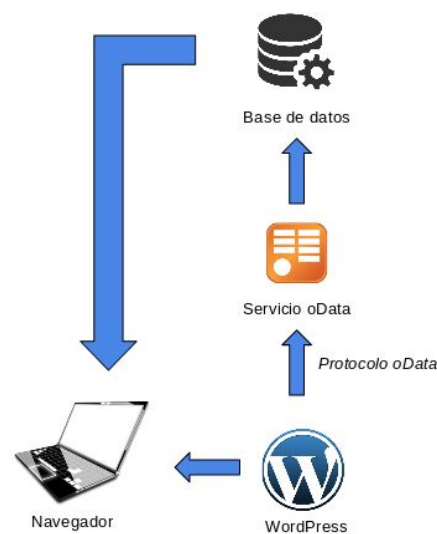


Fig. 1: Esquema de funcionamiento de *oData*

Se constata la dificultad que supone para la gran mayoría, el análisis de datos. Pese a la multitud de herramientas que existen en la actualidad, se hace difícil su uso por parte de profanos en la materia.

En este sentido, se propone utilizar WordPress como caso de uso, de cara a

<sup>23</sup> [www.rcommander.com](http://www.rcommander.com)

<sup>24</sup> <https://rapidminer.com>

<sup>25</sup> <http://www.odata.org/>

<sup>26</sup> <https://bitworking.org/projects/atom/rfc5023.html>

poder visualizar datos en este conocido CMS sin necesidad de tener conocimientos en la materia.

De esta forma se propone el acceso a la disciplina de análisis de datos para el gran público.

Previo confección del plugin para poder crear el caso de uso propuesto se realizó una investigación para verificar que no existía otro producto en el sector que se pudiera utilizar o readaptar. En este sentido se empezó la búsqueda en el repositorio oficial de WordPress, y los productos encontrados eran o demasiados antiguos para poder adaptarlos al funcionamiento actual de la herramienta o no se ajustaban a los requerimientos del caso de uso<sup>27 28</sup>. WordPress se actualiza dos veces al año, y al haber pasado tanto tiempo, la actualización de dichos plugins se hacía farragosa.

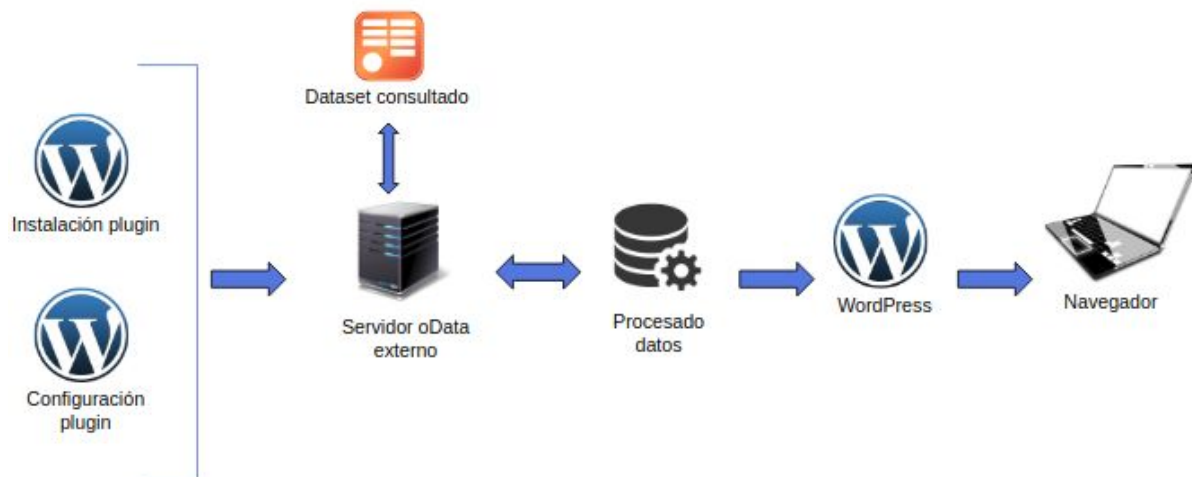


Fig. 2: Esquema de funcionamiento del plugin

En este caso de uso se propone, mediante una fuente de datos externos, representar la tabla con los datos necesarios dentro de un post de WordPress. Del mismo modo, se pueden filtrar los datos según sean las necesidades del usuario.

El usuario, después de instalar y configurar el plugin dentro de la instalación WordPress realiza una petición a la fuente de datos solicitada. Una vez la fuente (la cual funciona bajo oData y por tanto, expone los datos en esa tecnología) retorna los datos en formato

JSON, estos se devuelven a WordPress, dónde el plugin los clasifica y los ordena de forma que se puedan mostrar correctamente en el navegador del usuario según las opciones que este haya previamente fijado en el primer paso.

El funcionamiento básico del plugin se puede ver en la siguiente figura 2.

La entidad básica de información de WordPress es una entrada o *post*, dónde el usuario introduce la información que quiere mostrar. La herramienta que WordPress provee para que el usuario introduzca esa información dentro del sistema es el editor HTML TinyMCE<sup>29</sup>. Dicho editor incorpora una

serie de botones para poder realizar las funcionalidades típicas de un documento, cambiar fuente, color o alineación de texto, por nombrar algunos.

El plugin incorpora un nuevo botón de forma que, al clicar en él se abre una ventana en la que el usuario introduce la URL del dataset (en formato oData) y el número de entradas que quiere mostrar (si no especifica nada, se mostrarán todas). Por último, existe filtro predefinido, opción mediante la cual se puede realizar una consulta inicial prefiltrada.

Asimismo, el usuario puede modificar el estilo de la tabla: el color de la cabecera, el tamaño

<sup>27</sup> <https://es.wordpress.org/plugins/wp-opendata/>

<sup>28</sup>

<https://github.com/PeopleAndCode/Wordpress-OData-API>

<sup>29</sup> <https://www.tinymce.com>



de fuente y una ventana propia desde la cual se puede incorporar la consulta predeterminada en formato oData en caso de que la persona posee conocimientos en dicho metalenguaje.

Una vez rellenados estos datos, el usuario siempre está a tiempo de poder cambiar los valores editando el código que se ha generado en el editor.

En la interfaz visual del plugin, esto es la propia web, el usuario puede ir refinando su consulta mediante los filtros añadidos a tal efecto, de forma que pueda realizar consultas al dataset para poder llegar finalmente a obtener los datos que desee.

Además, el *plugin* se encuentra localizado, esto es listo para ser traducido a diferentes idiomas.

## 6. Conclusiones

Los datos abiertos se han convertido en una fuente de conocimiento cada vez más presente en la sociedad. Además de la presencia en múltiples disciplinas, se ha incorporado en las administraciones centrales contribuyendo así al empoderamiento de los ciudadanos, generando así mayor conocimiento por parte de estos de la sociedad en la que habitan.

La disciplina del análisis de datos se establece como primordial en la sociedad actual. Alrededor del análisis de datos tradicional, surgen nuevas aplicaciones por parte de las administraciones centrales de cara a intentar mejorar la sociedad, destinadas a una mejor toma de decisiones.

A la vista de la investigación realizada, se constata la dificultad de explotación de datos por parte de la sociedad en general, y de la necesidad de obtener una forma de analizar datos que no requiera de una formación avanzada en el campo. En este sentido, se ha desarrollado un caso de uso con una herramienta que permita empezar a extraer datos de fuentes externas para poderlos integrar en una de las herramientas de gestión de contenido más importante de la actualidad: WordPress.

Aunque el foco del caso de uso se ha centrado en poder obtener datos externos de

una fuente, a modo de trabajo futuro se podría contemplar la opción de combinar dichos datos para poder integrar datos de diferentes tablas oData. A efectos de desarrollo, dicha funcionalidad implicaría poder usar como filtro personalizado la función *\$links*. Podemos observar el funcionamiento de la misma en el apartado 3.2 de la referencia de oData<sup>30</sup>.

En una URL tipo [http://services.odata.org/OData/OData.svc/Categories\(1\)/\\$links/Products](http://services.odata.org/OData/OData.svc/Categories(1)/$links/Products), se enlazarían los productos relacionados con la categoría 1.

Otras posibles mejoras del *plugin* se podrían centrar en la obtención de un mayor control de los campos a mostrar y un aumento de las traducciones disponibles.

---

30

<http://www.odata.org/documentation/odata-version-2-0/uri-conventions/>

## 7. Referencias

- Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In *Learning analytics* (pp. 61-75). Springer New York.
- Bauer, F., & Kaltenböck, M. (2011). Linked open data: The essentials. *Edition mono/monochrom, Vienna*.
- Berners-Lee, T. (2006). Linked Data - Design Issues. <https://www.w3.org/DesignIssues/LinkedData.html>
- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. *Semantic services, interoperability and web applications: emerging concepts*, 205-227.
- Bradley, E. H., Curry, L. A., & Devers, K. J. (2007). Qualitative data analysis for health services research: developing taxonomy, themes, and theory. *Health services research*, 42(4), 1758-1772.
- Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly*, 36(4).
- De Cáceres, M., Font, X., García, R. & Oliva, F. (2003). Vegana, un paquete de programas para la gestión y análisis de datos ecológicos .
- 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.
- Hsieh, W. W., & Tang, B. (1998). Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bulletin of the American Meteorological Society*, 79(9), 1855-1870.
- Huijboom, N., & Van den Broek, T. (2011). Open data: an international comparison of strategies. *European journal of ePractice*, 12(1), 4-16.
- Janssen, M., Charalabidis, Y., & Zuidewijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information systems management*, 29(4), 258-268.
- Odom, M. D., & Sharda, R. (1990, June). A neural network model for bankruptcy prediction. In *Neural Networks, 1990., 1990 IJCNN International Joint Conference on* (pp. 163-168). IEEE.
- Pope, C., Ziebland, S., & Mays, N. (2000). Qualitative research in health care: analysing qualitative data. *BMJ: British Medical Journal*, 320(7227), 114.
- Sandoval-Almazán, R., Gil-García, J. R., Luna-Reyes, L. F., Luna, D. E., & Rojas-Romero, Y. (2012, October). Open government 2.0: citizen empowerment through open data, web and mobile apps. In *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance* (pp. 30-33). ACM.
- Tukey, J. W. (1962). The future of data analysis. *The annals of mathematical statistics*, 33(1), 1-67.
- Villena-Román, J., Cobos, A. L., & Cristóbal, J. C. G. (2014, July). TweetAlert: Semantic Analytics in Social Networks for Citizen Opinion Mining in the City of the Future. In *UMAP Workshops*.
- Weir, B. S. (1990). Genetic data analysis. Methods for discrete population genetic data. *Sinauer Associates, Inc. Publishers*.