

Aragón Open Data

**CKAN, piedra angular para la
gestión de un sistema de
datos abiertos**

Índice

1 Solución Open Source para la gestión de datos abiertos 3
1.1 Autor 7

1 Solución Open Source para la gestión de datos abiertos

CKAN como solución software de gestión de datos abiertos en un caso real para el portal Aragón Open Data

Aragón Open Data es el portal de datos abiertos del Gobierno de Aragón. El portal está administrado por la Dirección General de Administración Electrónica y Sociedad de la Información y tiene como misión ser el catálogo a partir del cual los ciudadanos y las empresas puedan acceder al conjunto de datos abiertos que publiquen tanto el Gobierno de Aragón, como el resto de instituciones del territorio aragonés que se quieran unir a la iniciativa.

El principal objetivo de un proyecto de datos abiertos es la generación de un impacto positivo y constructivo gracias a una serie de recursos de carácter político, con beneficios sociales o que ayuden a un crecimiento económico.

Open Data es una iniciativa a nivel mundial que persigue que toda la información, sobre todo a nivel gubernamental e instituciones públicas, estén disponibles para todos los ciudadanos y que puedan ser legibles y entendibles por la gran mayoría.

El Gobierno de Aragón, para este proyecto de datos abiertos, confió desde sus inicios en un sistema de gestión de datos abiertos llamado **CKAN**. Muchas veces se tiende a explicar esta herramienta haciendo alusión a WordPress como un gestor de contenido, en el que el contenido suelen ser páginas y publicaciones en un blog. Y en el caso de un gestor de datos abiertos, ese contenido son conjuntos de datos, los cuales a su vez, contienen diferentes recursos en formatos reutilizables.

Esta plataforma ayuda a administrar y publicar colecciones de datos. Y es utilizado por gobiernos locales, nacionales e internacionales, además de instituciones de investigación y otras organizaciones que recopilan una gran cantidad de datos. La extensión de su uso se ha globalizado de tal manera que se ha convertido en una herramienta estándar dentro del mundo de los datos abiertos.

El backend de CKAN está desarrollado íntegramente en Python, con un frontal propio en Javascript. Posee una capa de servicios en forma de API que potencia su uso, y existe la posibilidad de utilizar plugins base, que se detalla más adelante, para implementar un sistema de recolección de datos como el que se explicó en un [artículo anterior](#).

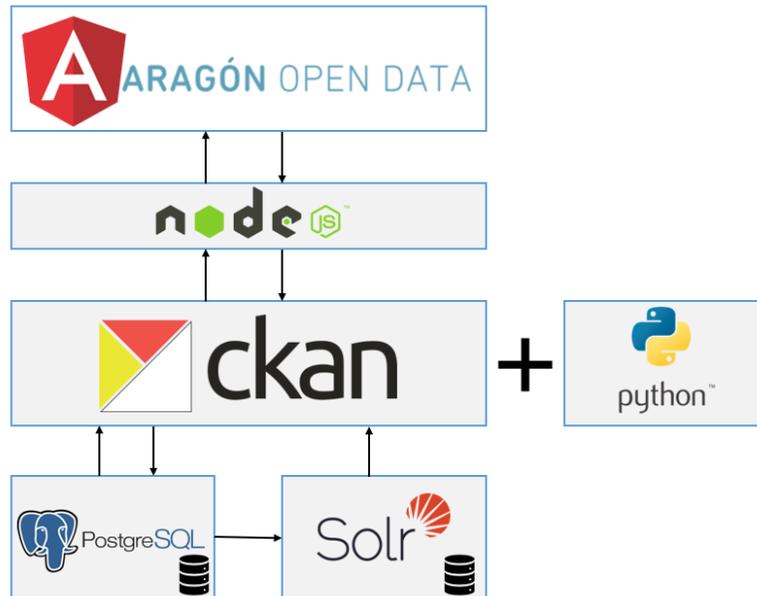
Además, gracias a la ya mencionada API pública, se exponen los datos al exterior facilitando la información a otras aplicaciones de terceros como apps móviles o como es el caso en Aragón Open Data, que utiliza este mecanismo para construir un frontal nuevo y más moderno con tecnología Angular.

Uno de los retos a los que nos enfrentamos durante este proyecto fue que toda la personalización de CKAN, páginas, funcionalidad extra u otros tipos de desarrollos, estaban implementados directamente en el núcleo de CKAN, lo que dificultó la migración a una versión superior, ya que las nuevas actualizaciones dejaron de ser compatibles. En este caso, la necesidad pasó por volver a desarrollar todo sobre la nueva versión, pero esta vez utilizando plugins y/o componentes externos al núcleo.

La solución escogida e implementada fue la de construir una aplicación cliente para evitar el acoplamiento de CKAN a la capa de visualización de los datos y al panel de administración para el equipo de Aragón Open Data. Con esto se facilitó la posibilidad de que el día de mañana se requiera instalar otro portal específico apuntando a la misma instancia de CKAN.

Esta aplicación cliente consta de una capa de Angular como frontal y una capa de servicio en NodeJS para gestionar de una manera más óptima todas las peticiones a la API de CKAN. Así extrajimos las páginas añadidas dentro del núcleo de CKAN a componentes Angular, al igual que las funcionalidades de backend, que se separaron entre la parte de servicios de NodeJS y una opción de plugins que añade funcionalidad a la plataforma, que son fácilmente migrables y que explicaremos más adelante en este artículo.

Para esto se propuso la siguiente arquitectura:



Como se puede observar en el diagrama anterior, CKAN se apoya en una base de datos **PostgreSQL**, donde se almacenan todos los conjuntos de datos, recursos y demás metadatos necesarios para el funcionamiento de la plataforma.

Por otro lado, CKAN hace uso de **Solr**, un motor de búsqueda que realiza su función sobre un determinado conjunto de documentos, ya sean textos, bases de datos, etc... Solr realiza un proceso de indexado mediante el cual incluye los términos de archivos específicos en un índice. De esta manera se agilizan las búsquedas y las relaciones entre los conjuntos de datos.

Además, Solr es una de las herramientas de indexación más utilizadas en el mundo de Open Data, junto a **ElasticSearch**, el cual también se está utilizando en el ecosistema de Aragón Open Data, más en concreto en el servicio Open Analytics Data.

Volviendo al núcleo de la tecnología del stack presentado, CKAN ofrece dos opciones principales para añadir funcionalidad a la plataforma de datos abiertos, ambas soluciones son a través de plugins escritos en Python:

Extensiones core: algunas de estas extensiones vienen empaquetadas en CKAN y no necesitan ser instaladas para usarlas. Es suficiente con activarlas desde el fichero de configuración de CKAN. Algunos de estos componentes extras son: datastore, multilingual, stats y data viewer.

- La extensión **datastore** proporciona una base de datos ad hoc para el almacenamiento de datos estructurados de los recursos de CKAN.
- La extensión **multilingual** permite al administrador introducir traducciones en la base de datos para los términos que pueden aparecer en el contenido de los conjuntos de datos, grupos o etiquetas creadas por los usuarios.
- La extensión **stats** analiza la base de datos y ofrece varias tablas y gráficos con estadísticas del portal, por ejemplo: número total de datasets, los datasets más vistos, los usuarios que poseen más datasets, etc...
- La extensión **data viewer** muestra una previsualización de los recursos de cada conjunto de datos, ya sea en forma de tabla, gráfico o mapa. Para que esta extensión funcione de manera

correcta, la calidad y la estructura de los datos deben ser muy estricta y cumplir prácticamente todas las **reglas** que conlleva abrir un recurso como formato reutilizable.

Extensiones externas: son extensiones que no vienen instaladas en el núcleo de CKAN, pero pueden ser descargadas e instaladas de forma independiente. Se pueden buscar en <https://extensions.ckan.org/>. Además, podemos crear nuestras propias extensiones, subirlas a un repositorio de código, por ejemplo github, e instalar nuestra extensión como cualquier otra. Esta opción te permite añadir funcionalidad al núcleo de CKAN sin modificarlo y así, en un futuro, si es necesario realizar una migración de la plataforma, bastará con instalar la nueva versión de CKAN e instalar de nuevo cada una de las extensiones.

Por ejemplo, en el portal de Aragón Open Data se ha trabajado en distintas funcionalidades que se han tenido que desarrollar a medida, como los distintos harvesters de conjuntos de datos del Instituto Aragonés de Estadística o de la Universidad de Zaragoza. Estos recolectores han sido desarrollados como plugins externos e instalados posteriormente en la plataforma.

En los distintos eventos de datos abiertos que se realizan durante el año, siempre se hace hincapié en la importancia de la reutilización de los datos que ofrecen estas plataformas. Los datos están para todos, tanto empresas como ciudadanos. Y, además, como suelen ser datos de carácter público, procedentes de instituciones u organizaciones públicas, pueden resultar muy interesantes y enriquecedores.

En Aragón Open Data se ofrecen diversos servicios, los cuales ya están reutilizando estos datos abiertos, y que ofrecen al ciudadano la posibilidad de entender ficheros CSV o JSON de una forma más visible. Por ejemplo, uno de los servicios más visitados es **Presupuestos de Aragón**, donde se muestran de un modo claro cómo se distribuye el presupuesto del Gobierno de Aragón.



Es importante dar a conocer el último servicio que se ha añadido a la plataforma, en el que también hemos trabajado como proyecto independiente al del portal de Aragón Open Data, y que no es otro que el **Open Data Focus**. Un servicio que te permite crear historias y gráficos reutilizando los datos abiertos disponibles.



Esto no debería quedar en servicios que propone u ofrece el propio portal de Aragón Open Data, sino que la idea es que cualquier desarrollador pudiera ser capaz de construir una aplicación que se nutra de estos datos abiertos para convertirlos en valor para el resto de la ciudadanía.

En resumen, el conjunto de componentes Angular, NodeJS, CKAN, PostgreSQL y Solr se puede utilizar para el desarrollo de soluciones de datos abiertos, en los que se desea exponer un volumen de recursos con numerosas relaciones, en un amplio abanico de formatos reutilizables y que su a vez tiene que estar disponible para consumir datos diariamente y que sean consumidos por terceros, mediante APIs.

Podría encajar como solución a distintos problemas muy relacionados con la cantidad de datos y su necesidad para indexar búsquedas, sin embargo, si el proyecto estuviera totalmente fuera del contexto de los datos abiertos, el componente de CKAN se debería extirpar y perdería sentido todo el contexto del stack mostrado aquí.

La dificultad de la implementación del stack reside en si es una instalación de cero o no. En el caso de que sea de cero, la única dificultad estaría fundamentada en la diferencia de tecnologías, ya que estaríamos hablando de Typescript, Javascript, Python, PostgreSQL y Solr.

Si el proyecto es de datos abiertos es una buena decisión trabajar con el stack entero. La facilidad que ofrece Angular para modificar o añadir nuevos componentes visuales, NodeJS para adaptar los servicios que se vayan requiriendo y CKAN para gestionar datasets y añadir funcionalidad de datos abiertos a través de sus plugins es una buena opción para utilizar esta arquitectura Open Data.

1.1 Autor



Gabriel Alcober Fuertes

Equipo de Aragón Open Data | Delivery Lead en DXD

Gabriel inició su carrera profesional en 2011, y durante este tiempo ha participado en proyectos con gran diversidad de tecnologías. Ha trabajado en proyectos Salesforce y Java, y desde 2017 ha trabajado en soluciones enfocadas a datos abiertos y web semántica.