



CÓMO PUBLICAR DATOS ABIERTOS DE MANERA RÁPIDA Y SENCILLA (CON CKAN)

El objetivo que se persigue con los **datos abiertos** (open data) es aparentemente sencillo: la publicación de esos datos de forma que puedan ser reutilizados y que todo el mundo se pueda beneficiar de ellos.

Esta guía está destinada a todos aquellos que deseen empezar a compartir sus datos pero dispongan de recursos limitados para poder hacerlo. El objetivo de la guía es servir de ayuda y orientación para que cualquiera pueda abrir fácilmente sus datos, aunque no tenga grandes conocimientos sobre estrategias de apertura de la información o las tecnologías asociadas a ellas¹.

En la primera parte se establecen una serie de sencillas pautas que servirán de guía a la hora de localizar y preparar los datos para su publicación. A continuación se indicará como publicar los datos en la Web a través de un catálogo y de la forma más simple posible. Para ello utilizaremos como ejemplo [CKAN](#), una plataforma de publicación de datos de código abierto disponible gratuitamente, aunque las recomendaciones generales de esta guía serían de aplicación también con cualquier otra plataforma que se desee utilizar.

SELECCIÓN DE LOS DATOS QUE SE VAN A PUBLICAR

El objetivo final de cualquier iniciativa de datos abiertos siempre será conseguir **publicar todos los datos disponibles**. Sin embargo, cuanto más limitados sean el tiempo y los recursos disponibles, más importante será determinar cuál es la información más valiosa desde el punto de vista de la *reutilización* para poder dar mayor prioridad a su publicación.

Criterios

La priorización debería adecuarse a las particularidades de cada caso. Los gestores de la información serán los mejores conocedores de los datos que manejan y de su potencial audiencia, pero algunos criterios generales que pueden servir de guía general son:

- **Exigencias de la [normativa](#) existente**, siguiendo las recomendaciones y legislación aplicables a nivel europeo, nacional o regional.
- **Importancia de los datos para la sociedad**, teniendo en cuenta el posible impacto y utilidad de los datos para el conjunto de la sociedad.
- **Importancia de los datos para la economía**, valorando el potencial de los datos para generar nuevas líneas de investigación o negocio.

¹ La guía ha sido diseñada para ser útil incluso cuando no se tengan conocimientos técnicos, no obstante será mucho más fácil seguirla si se cuentan ya con ciertos conocimientos básicos.

- **Actualidad de los datos**, ya que en general cuanto más actualizados mayor potencial, aunque los datos históricos también pueden tener un gran valor.
- **Cantidad de datos**, dado que cuanto mayor sea el volumen más valor añadido se podría aportar con el tratamiento posterior de la información.
- **Nivel de detalle**, puesto que cuanto más detallados sean los datos mayores serán también las posibilidades de explotación.

Temáticas

En cuanto a las temáticas más demandadas, existen varios criterios de referencia promovidos por organismos oficiales de relevancia en la materia tales como:

- Los resultados del estudio [MEPSIR](#) de la Unión Europea;
- Las materias que generan negocio por parte del sector infomediario en España, según el [análisis del sector realizado por ONTSI](#);
- Los datos comunes a las distintas iniciativas open data en España;
- Los [conjuntos de datos prioritario definidos por el G8](#);
- Las [recomendaciones](#) de la Comisión Europea.
- La [disponibilidad mundial de datos](#) analizada por el [Open Data Barometer](#);
- La [clasificación de sectores temáticos](#) utilizada por datos.gob.es.

En un primer análisis general se podrían sugerir como temáticas de alto interés los pertenecientes a los siguientes ámbitos de información:

Ámbitos de información	
Negocio y Comercio	Comercio y exportación
	Empresas y actividades comerciales
	Construcción
	Contrataciones públicas
Medio Ambiente	Meteorología
	Sostenibilidad
	Contaminación
	Recursos naturales
Geografía	Información topográfica
	Callejero
	Información catastral
Sociedad	Cultura
	Sanidad y bienestar
	Formación y Empleo
	Ocio y turismo
	Seguridad
	Educación
	Censo y demografía
	Vivienda
Transporte	Tráfico
	Transporte público
	Registro de vehículos



Ámbitos de información	
Legal	Justicia
	Registros públicos
	Legislación
Administración pública	Organigrama
	Servicios públicos
	Transparencia
	Infraestructuras
	Elecciones
	Presupuestos y gasto

En cualquier caso, cada publicador debe también completar y adaptar su propia lista de criterios e intereses de acuerdo a su actividad.

Restricciones

Aunque la norma general consiste simplemente en **publicar cuanto más mejor**, existen también limitaciones a la hora de publicar datos que se deben respetar:

- Datos restringidos por la ley protección de datos por su **carácter personal**;
- Datos sometidos al **secreto estadístico**;
- Datos sujetos a la **confidencialidad comercial**;
- Datos que afecten a la protección de la **seguridad pública**;
- Datos sujetos al **derecho de propiedad intelectual o industrial**;

En cualquiera de estos casos la información restringida debería ser filtrada y eliminada **previamente a su publicación**.

LICENCIAS PARA FACILITAR LA REUTILIZACIÓN

Es fundamental que cualquier dato publicado incluya **información explícita sobre las condiciones de uso** de esa información para dejar claro a los potenciales usuarios cuáles son las posibilidades de reutilización de esos datos. Para hacerlo de forma simple contamos con dos opciones alternativas:

Establecer unas condiciones de uso simples

En este caso indicaremos que los datos están plenamente disponibles para su reutilización bajo una serie de condiciones mínimas, para las cuales podemos guiarnos por ejemplo por las indicadas en el [Real Decreto 1495/2011, de 24 de octubre](#):

1. Está prohibido **desnaturalizar** el sentido de los datos.
2. Debe citarse la **fuentes** de los datos.
3. Debe mencionarse la fecha de la **última actualización**.



4. No se podrá indicar, insinuar o sugerir que los titulares de los datos participan, patrocinan o apoyan la reutilización que se lleve a cabo.
5. Deben conservarse y no alterarse los metadatos sobre la actualización y las condiciones de reutilización aplicables.

Adoptar una licencia predefinida

En este otro caso simplemente indicaríamos que los datos están disponibles para su reutilización bajo las condiciones establecidas por alguna de las [licencias compatibles con los principios de los datos abiertos](#). Estas licencias podrán a lo sumo establecer restricciones relacionadas con la atribución de la fuente original y la obligatoriedad de compartir cualquier trabajo realizado con los datos bajo las mismas condiciones. Podremos elegir por ejemplo entre las siguientes opciones:

1. Licencias de **dominio público**, que nos permiten compartir los datos sin ningún tipo de restricción:
 - [Creative Commons Zero](#) (CC Zero)
 - [Public Domain Dedication and License](#) (PDDL 1.0).
2. Licencias que únicamente obligan a **atribuir** la fuente original de los datos:
 - [Creative Commons-Reconocimiento](#) (CC BY 4.0)
 - [Open Data Commons Attribution License](#) 1.0 (ODC-By 1.0).
3. Licencias que requieren **atribución** de la fuente original y compartir los resultados bajo las **mismas condiciones**:
 - [Creative Commons-Reconocimiento Compartir Igual](#) (CC BY-SA 4.0)
 - [Open Data Commons Open Database License](#) (ODbL 1.0).

Importante también señalar que aunque no se quiera poner ningún tipo de restricción sobre la reutilización de los datos hay que indicarlo explícitamente mediante unas condiciones de uso o licencia de dominio público, ya que de otro modo tendríamos dos problemas:

- Los potenciales reutilizadores no sabrían cuáles son las condiciones aplicables en la práctica por no tener ninguna indicación sobre ello.
- Podrían seguir aplicándose algunos derechos *por defecto* (por ejemplo relacionados con los derechos de actor o de propiedad de las bases de datos) al no haber renunciado explícitamente a ellos mediante la declaración de dominio público u otro tipo de condiciones.

PREPARACIÓN DE LOS DATOS

Uno de los principales requisitos para la reutilización es que el **procesamiento automático de los datos** sea posible, ya que esta es la única manera de trabajar con grandes cantidades de datos de forma eficiente. Para ello será necesario que la

información esté publicada en los **formatos adecuados** y con el suficiente nivel de calidad y detalle para que los datos sean aprovechables.

Formatos

Algunos criterios a seguir a la hora de seleccionar los formatos adecuados son:

1. Formatos **estructurados** que siguen **modelos de datos estandarizados**.
2. Formatos directamente **legibles por las máquinas**.
3. Los formatos **más demandados** entre nuestros potenciales reutilizadores.
4. Formatos **abiertos** que permitan un uso **sin restricciones** para ampliar el público potencial.
5. Formatos que permitan publicar **metainformación** incorporada.
6. Formatos compatibles con lo establecido en la [Norma Técnica de Interoperabilidad del Catálogo de estándares](#).

Ejemplos de algunos formatos frecuentemente utilizados para la apertura de la información son:

Formato	Función
CSV (TSV)	Para representar datos en forma de tabla, donde las columnas se separan por comas (o tabulaciones) y las filas por saltos de línea.
ODS / XLS	Formatos (abierto en el caso de ODS y propietario de Microsoft en el caso de XLS) de hoja de cálculo que muestran la información en celdas organizadas en filas y columnas. Cada celda contiene datos o una fórmula con referencias relativas, absolutas o mixtas a otras celdas.
GML	Vocabulario XML para el modelaje, transporte y almacenamiento de información geográfica.
KML	Vocabulario XML para representar datos geográficos en tres dimensiones. Suelen distribuirse comprimidos en formato KMZ.
JSON	Notación muy popular para el intercambio de datos basada en una simplificación de la notación de JavaScript.
RDF	Modelo de representación para la descripción de recursos siguiendo los fundamentos de la Web Semántica.
RSS	Vocabulario XML para facilitar la distribución de contenidos y actualizaciones.
SHP	Formato propietario de datos espaciales que se ha convertido en un estándar <i>de-facto</i> y sirve para almacenar tanto la localización de los elementos geográficos como los atributos asociados a ellos.
WMS	Formato de servicio web para producir mapas de datos referenciados espacialmente, de forma dinámica a partir de información geográfica.

Formato	Función
XML	Lenguaje general de etiquetado que define un conjunto de reglas para la codificación de documentos a través de modelos específicos legibles y procesables.

Así mismo, y siempre que se pueda, es recomendable también ofrecer los datos en varios **formatos distintos simultáneamente** para así intentar dar respuesta a las necesidades del mayor número de *reutilizadores* posible.

Nivel de detalle

Es también importante **respetar el nivel de detalle original de los datos**, tal y como se generan en su origen, evitando cualquier modificación y alteración previa a su publicación para que cada *reutilizador* pueda luego realizar el tratamiento de la información que más le convenga.

La única excepción a esta regla será el filtrado que pueda ser necesario realizar debido a las cuestiones de **privacidad, confidencialidad o seguridad** que se han citado anteriormente.

Preparación de los metadatos

El último paso previo a la publicación de los datos consistirá en preparar los **metadatos** que los acompañarán proporcionando información adicional para facilitar su localización, clasificación y reutilización. Es muy importante no sólo proporcionar el mayor número de metadatos posible, sino hacerlo también de la forma correcta. Podemos dividir esos metadatos en:

- **Obligatorios:** aquellos que, por su especial relevancia y también por cumplimiento con la [legislación vigente](#), será obligado proporcionar.
- **Recomendados:** aquellos que, si bien no serían obligatorios para cumplir con la legislación vigente, siguen siendo relevantes.
- **Opcionales:** aquellos que, aún siendo igualmente recomendables, puede que no siempre estén disponibles.

Por otro lado, a su vez los metadatos servirán a veces para describir los **conjuntos de datos** y otras para describir los distintos **recursos** (o archivos) que los componen. A continuación se muestra una tabla resumen de los metadatos que podemos proporcionar:

TABLA RESUMEN METADATOS		
CONJUNTOS DE DATOS		
OBLIGATORIOS	RECOMENDADOS	OPCIONALES
Nombre	Etiqueta(s)	Fecha de creación
Descripción	Condiciones de uso / licencia	Fecha última actualización

Temática (s)	Frecuencia actualización	
Organismo	Idioma(s)	
	Cobertura geográfica	
	Cobertura temporal	
	Vigencia	
	Recursos relacionados	
	Normativa	
<i>DISTRIBUCIÓN (archivos o enlaces con los datos)</i>		
OBLIGATORIOS	RECOMENDADOS	OPCIONALES
URL de acceso	Nombre	Tamaño
Formato	Información sobre formato	

Para una descripción más detallada sobre cada uno de estos metadatos podemos consultar el Anexo III de la [NTI de Reutilización de recursos de información](#) y su [guía de aplicación](#).

PUBLICAR LOS DATOS (con CKAN)

Para una gestión óptima de los datos de cara a su reutilización es necesario establecer una recopilación de los mismos en un punto de acceso común, que generalmente denominamos **catálogo de datos**, lo que permitirá una organización más eficaz y un acceso más fácil por parte de los potenciales *reutilizadores*.

En esta guía utilizaremos como ejemplo [CKAN](#) porque es una plataforma de código abierto ampliamente utilizada y que está disponible gratuitamente, aunque existen también muchas otras opciones disponibles.

CKAN es una herramienta concebida para facilitar la publicación de catálogos de datos². Se podría comparar con un sistema de gestión de contenido tradicional (CMS), pero orientado a la publicación de datos en lugar de otro tipo de contenidos. Nos permitirá publicar y administrar colecciones de datos con muy pocos recursos y, una vez publicados, los usuarios podrán explorarlos e incluso pre-visualizarlos a través de mapas, gráficas y tablas.

Requisitos previos

CKAN puede instalarse en cualquier máquina con una capacidad normal. Como ejemplo de requisitos mínimos para una instalación básica podríamos establecer una máquina con doble núcleo, 4 GB de RAM y 60 GB de espacio disponible en disco. En cuanto a los requisitos de software, se establecen los siguientes:

- Sistema Operativo:
 - Ubuntu 12.04 64bit (recomendado) u otra distribución de Linux.

² Una versión de demostración de CKAN está disponible en <http://demo.ckan.org/es/>

- Servidor Web:
 - Apache/Nginx
 - Python 2.6/2.7x
- Servidor de base de datos:
 - Tomcat/Jetty
 - Solr
 - Postgress 8.4+

Instalación básica

Si nuestros conocimientos técnicos son limitados, la forma más rápida y sencilla de empezar a trabajar con nuestro catálogo de datos puede ser utilizar una versión pre-instalada de CKAN en un servicio en la nube como [Amazon Web Services \(AWS\)](#) con tan solo unos pocos pasos. Esta opción es sin duda la más simple, pero hay que tener en cuenta también que **conllevaría un coste por uso de los servicios de AWS desde 70\$ mensuales**³ (para un uso de 24 horas al día/30 días al mes y la configuración básica que vamos a comentar).

Para ello una vez hayamos creado nuestra cuenta de usuario en AWS⁴, accederemos a la [versión pre-instalada](#), elegiremos la región donde queremos desplegar la aplicación (*EU West – Ireland* para alojamiento de los datos en Europa) [Paso 1] y seleccionaremos la opción “Continuar” (*Continue*) [Paso 2]



A continuación escogeremos el tipo de servidor (*m1.medium* debería ser suficiente para empezar) [Paso 3] y tendremos que generar una nueva clave⁵ (*key pair*) para poder acceder luego al servidor. Usaremos entonces la opción “*Visit the Amazon EC2 Console*” [Paso 4] que abrirá la Consola de gestión en una ventana nueva.

³ Para un cálculo personalizado utilizar <http://calculator.s3.amazonaws.com/index.html>

⁴ Para crear la cuenta será necesario proporcionar un número de tarjeta de crédito.

⁵ En el caso de que ya seamos usuarios de AWS y tengamos alguna clave ya generada podremos simplemente seleccionarla en lugar de crear una nueva.

Launch on EC2:

CKAN delivered by Link Digital

Ya en la consola usaremos la opción “Create Key Pair” [Paso 5], introduciremos un nombre para la clave [Paso 6] y la generaremos con la opción “Create” [Paso 7].

Se generará un archivo con el nombre de clave que hayamos definido y la extensión *.pem* que deberemos guardar y reservar en una localización segura. Volveremos entonces a la página anterior y la recargaremos para comprobar que la clave ya está asignada [Paso 8] y podremos arrancar finalmente nuestro nuevo CKAN utilizando la opción “Launch with 1-Click” [Paso 9].

Launch on EC2:

CKAN delivered by Link Digital

Inmediatamente aparecerá un mensaje indicándonos que la *instancia* de CKAN que acabamos de crear se está inicializando y en unos pocos minutos estará lista para

ser utilizada. Podemos comprobar su estado en todo momento a través del acceso al panel de control [Paso 10].

An instance of this software is now deploying on EC2.

- If you would like to check the progress of this deployment, go to the [AWS Management Console](#). **Paso 10**
- The software will be ready in 2-3 minutes.

Usage Instructions
After the instance is started, you will be able to use the CKAN system via a web browser to upload datasets. In order to access all of the functions of the system, you will need to create a sysadmin account via SSH by following the instructions from... [Show more](#)

Software Installation Details

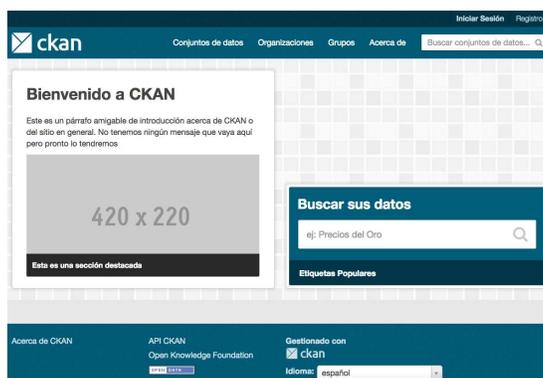
Product	CKAN delivered by Link Digital
Version	2.2*, released 11/20/2014
Region	EU West (Ireland)
EC2 Instance Type	m1.medium
VPC	vpc-65e03c00
Subnet	subnet-19e13640
Security Group	CKAN delivered by Link Digital-2--AutogenByAWSMP
Key Pair	KP-Europe-CKAN

Ya en el panel de control podremos ver cuándo CKAN está ya en marcha (el estado pasará a ser *“running”*) [Paso 11] y verificar que funciona correctamente accediendo a través de un navegador web utilizando de la IP pública que tiene asignada [Paso 12].



The screenshot shows the AWS Management Console interface. At the top, there are buttons for 'Launch Instance', 'Connect', and 'Actions'. Below that, there are two tabs: 'Paso 11' and 'Paso 12'. The 'Paso 11' tab is active, showing a table of instances. The first instance is highlighted, and its 'Instance State' is 'running'. The 'Public IP' column shows '54.72.250.114'. The 'Paso 12' tab is also visible, showing the 'Public IP' and 'Key Name' for the instance.

Si todo ha salido correctamente esto es lo que veremos en nuestro navegador al introducir la IP asignada en la barra de direcciones.



The screenshot shows the CKAN website homepage. The header includes the CKAN logo and navigation links for 'Conjuntos de datos', 'Organizaciones', 'Grupos', and 'Acerca de'. There is a search bar for datasets. The main content area features a 'Bienvenido a CKAN' message and a 'Buscar sus datos' section with a search input field. The footer contains information about CKAN, the Open Knowledge Foundation, and the language setting (español).

Ahora podríamos ya registrarnos directamente y empezar a utilizar CKAN como usuarios para publicar datos.

Instalación avanzada

Si contamos con unos conocimientos técnicos más amplios y queremos optar por una instalación más personalizada, la más cómoda entre las [opciones disponibles](#) sería utilizar la versión empaquetada para [servidores Ubuntu 12.04, 64-bit](#). Aquí contaríamos con dos alternativas respecto al *hardware* a utilizar:



1. Nuevamente podríamos optar por utilizar un servicio en la nube como AWS, aunque en esta ocasión partiríamos simplemente de una [preinstalación de Ubuntu](#) y necesitaríamos crear un grupo de seguridad (*security group*) con permisos 0.0.0.0 para los puertos 22, 80 y 5000. Contaríamos además con la ventaja añadida de poder seleccionar servidores más asequibles para realizar nuestras pruebas (*m1.small por unos 35\$ mensuales*⁶) y además podríamos optar también por utilizar una [cuenta de prueba gratuita durante 1 año](#) disponible para los servidores más básicos⁷ (*t1.micro*).
2. La otra posible alternativa en este caso sería realizar la instalación en algún servidor propio donde pudiéramos instalar o tuviéramos ya instalada la versión de Ubuntu correspondiente.

En cualquier caso, una vez solucionadas las necesidades de hardware procederíamos a instalar CKAN en [cuatro sencillos pasos](#) desde la línea de comandos de nuestro Ubuntu ya instalado:

1. Actualizar el índice de paquetes de Ubuntu:
`sudo apt-get update` □
2. Instalar los paquetes adicionales que necesita CKAN:
`sudo apt-get install -y nginx apache2 libapache2-mod-wsgi libpq5` □
3. Descargar el paquete de CKAN:
`wget http://packaging.ckan.org/python-ckan_2.2_amd64.deb` □
4. Instalar el paquete de CKAN:
`sudo dpkg -i python-ckan_2.2_amd64.deb` □

Además debemos asegurarnos de que tanto la base de datos PostgreSQL como Solr están correctamente instalados y configurados. Para ello:

1. Instalaremos PostgreSQL y Solr:
`sudo apt-get install -y postgresql solr-jetty` □
2. Seguiremos las instrucciones para configurar Solr y la [base de datos](#).
3. Editaremos la opción [sqlalchemy.url](#) del fichero `/etc/ckan/default/production.ini` para incluir el nombre de la base de datos que acabamos de configurar, el usuario y la contraseña.
4. Inicializaremos la base de datos:
`sudo ckan db init` □

Finalmente reiniciamos Apache y Ngix

```
sudo service apache2 restart □  
sudo service nginx restart □
```

⁶ Para un cálculo personalizado utilizar <http://calculator.s3.amazonaws.com/index.html>

⁷ Los servidores básicos tienen unas características muy limitadas y es probable que no sirvan para cualquier cosa que no sean pruebas.



Si todos los pasos se han ejecutado correctamente ahora ya podremos acceder a nuestro CKAN a través de <http://localhost> o el DNS que le hayamos asignado.

Administración

Una vez tenemos CKAN instalado y funcionando, el siguiente paso será crear un usuario administrador para poder realizar la configuración básica de nuestro catálogo. **Si estamos usando AWS** lo primero será conectarse a nuestra máquina remota siguiendo estos pasos desde la línea de comandos:

1. Establecer los permisos adecuados para nuestro fichero con la clave que ya hemos creado previamente (Paso 7):

```
chmod 400 nuestro_fichero_clave.pem □
```

2. Conectarnos como administrador (*ec2-user*) a nuestra instancia de CKAN a través de la IP pública que le ha sido asignada previamente (Paso 12):

```
ssh -i nuestro_fichero_clave.pem ec2-user@IP.IP.IP □
```

Después, **tanto si estamos usando AWS o un servidor propio**, seguiremos los siguientes pasos nuevamente a través de la línea de comandos de Ubuntu:

1. Activar nuestro entorno virtual:

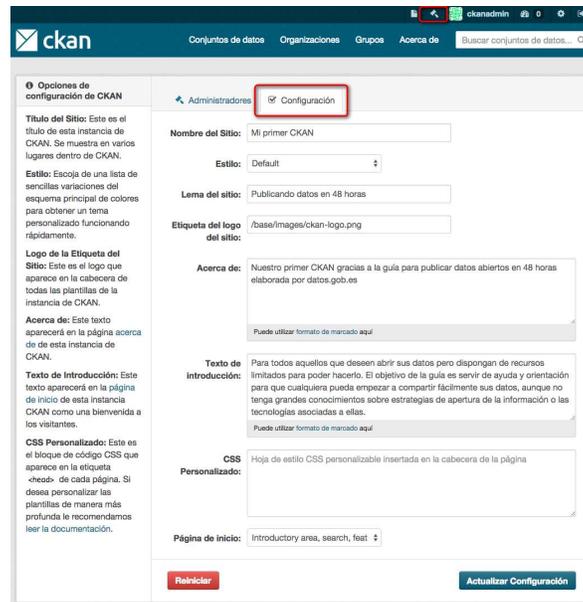
```
./usr/lib/ckan/default/bin/activate □  
cd /usr/lib/ckan/default/src/ckan □
```

2. Crear un usuario administrador con nombre “*ckanadmin*”:

```
paster sysadmin add ckanadmin -c /etc/ckan/default/production.ini □
```

3. Introducir la contraseña para el usuario administrador.

Con esto ya podremos acceder a nuestro CKAN como administradores y cambiar la su configuración básica para adaptarla a nuestras preferencias. Para ello nos identificaremos a través de la opción “Iniciar Sesión” del menú superior y utilizaremos la opción con el icono del “martillo” - también en la parte superior - para acceder al menú de administración donde desde la opción “Configuración” podremos cambiar aspectos básicos como el nombre del sitio, el estilo de presentación, el lema y el tema de presentación.



Ahora ya tenemos nuestro propio catálogo preparado y listo para empezar a compartir datos.

Características avanzadas

Para una personalización más avanzada de la apariencia del catálogo deberemos acudir a las recomendaciones de la [guía de personalización](#) de CKAN y para conocer todas las posibilidades de configuración tenemos la [guía avanzada de administración](#).

Además, CKAN cuenta con un amplio [catálogo de extensiones](#) que nos permitirá ampliar considerablemente su funcionalidad de una forma limpia y sencilla. Incluso los usuarios más avanzados podrán crear sus propias extensiones para añadir nueva funcionalidad.

También disponemos de una completa [interfaz de programación](#) (API) que facilitará la creación de aplicaciones y servicios a partir de los datos publicados.

Compatibilidad con la NTI

En este punto es importante también llamar especialmente la atención sobre la [Norma Técnica de Interoperabilidad para la Reutilización de recursos de información](#), ya que es la normativa nacional de referencia a la hora de garantizar la interoperabilidad y facilidad de uso de los datos que publicamos para su reutilización.

Existe el inconveniente de que CKAN presenta varios [problemas de compatibilidad con la NTI](#)⁸, por lo que no sería posible cumplir completamente dicha norma sólo con una instalación de CKAN por defecto como la que hemos realizado en esta guía.

⁸ En esta comparativa se incluye un análisis del nivel de implementación de la NTI en la versión 2.2a de CKAN en el que se ve claramente que la compatibilidad es baja.

Para tratar de paliar estos problemas en la medida de lo posible y con el mínimo esfuerzo podríamos hacer uso de las mejoras que otros ya han realizado previamente, como por ejemplo:

- Instalar la extensión [CKAN+DCAT](#) que mejora considerablemente el soporte de [DCAT](#) en la plataforma (estándar de metadatos del W3C para facilitar la interoperabilidad de catálogos en el que se basa la NTI) .
- Reutilizar el código de alguna de las extensiones y mejoras realizadas por otras iniciativas previas de datos abiertos en nuestro país para mejorar la compatibilidad con la NTI y facilitar la *federación* de los datos de nuestra plataforma con el catálogo nacional de datos [datos.gob.es](#):
 - o [Extensión de Opendata Cáceres](#) para la federación con datos.gob.es
 - o [Extensión del portal de datos abiertos del Ayuntamiento de Málaga](#).
 - o [Personalización de Aragón Open Data](#).
- [Personalizar los formularios](#) de metadatos que se utilizan para introducir la información sobre los conjuntos de datos y poder así adaptarlos a las necesidades de la norma.

Carga de datos

Ahora que ya tenemos todo preparado, tan sólo queda empezar a añadir datos a nuestro catálogo. Lo primero que debemos hacer es registrarnos con nuestro usuario, seleccionar la opción “Conjuntos de Datos” del menú superior y luego la opción “Agrega conjunto de datos”.



A continuación podemos empezar a crear nuestro primer conjunto de datos para compartir añadiendo la información correspondiente a los metadatos del “Título”, la “Descripción”, las “Etiquetas” y la “Licencia”.

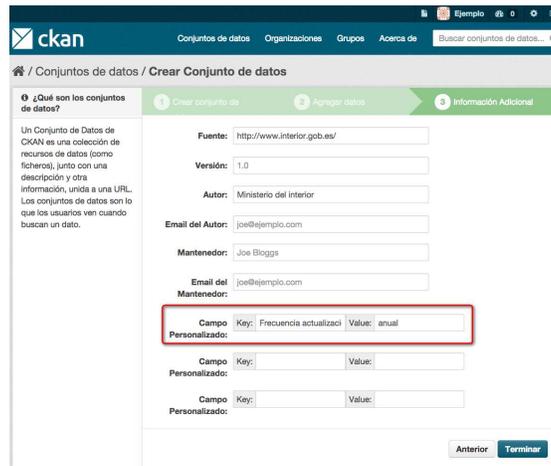
Una vez hayamos terminado accederemos al siguiente paso utilizando la opción “Siguiente: Agregar Datos” de la esquina inferior derecha. Ahora podremos ya elegir entre “Subir” directamente un fichero con los datos o añadir un “Enlace” directo a los datos si están ya disponibles en una dirección web.

Si elegimos por ejemplo añadir un “Enlace” simplemente tendremos que incluir la dirección web del sitio donde estén alojados los datos.

URL:

Por último completaremos con el “Nombre” y la “Descripción”, así como la extensión del formato en el que están guardados los datos (CSV, JSON, XML, etc.) Una vez hemos finalizado podemos elegir entre guardar estos datos y “Agregar” más datos a este mismo conjunto o terminar de añadir datos y pasar a proporcionar algo más de “Información adicional”.

Cuando hemos terminado de añadir todos los datos tan solo nos quedará introducir algo más de información adicional, como la “Fuente” original de los datos, la “Versión” el “Autor”, el “email” de contacto del autor, etc. También podremos añadir otros “Campos Personalizados” que podemos aprovechar para incluir otros metadatos que necesitemos (por ejemplo porque así lo indique la normativa) y no estén recogidos entre los anteriores. En este caso deberemos añadir tanto el nombre del metadato (Key) como su valor (Value).



ckan / Conjuntos de datos / Crear Conjunto de datos

1 Crear conjunto de datos 2 Agregar datos 3 Información Adicional

¿Qué son los conjuntos de datos?

Un Conjunto de Datos de CKAN es una colección de recursos de datos (como ficheros), junto con una descripción y otra información, unida a una URL. Los conjuntos de datos son lo que los usuarios ven cuando buscan un dato.

Fuente:

Versión:

Autor:

Email del Autor:

Mantenedor:

Email del Mantenedor:

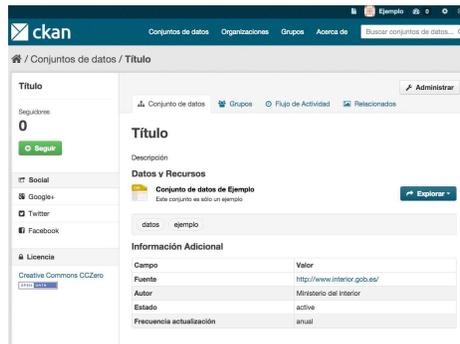
Campo Personalizado: **Key:** Frecuencia actualizaci **Value:** anual

Campo Personalizado: **Key:** **Value:**

Campo Personalizado: **Key:** **Value:**

Anterior Terminar

Este mismo proceso lo repetiremos tantas veces como sea necesario hasta que hayamos introducido todos los conjuntos de datos que tengamos previsto. A partir de este momento los datos estarán también disponibles a través del catálogo para todo aquel que quiera usarlos y experimentar con las distintas opciones de clasificación y visualización con las que cuenta la plataforma.



ckan / Conjuntos de datos / Título

Título

Regulaciones: 0

Social: Social, Google+, Twitter, Facebook

Licencia: Creative Commons CCZero

Administrar

Conjunto de datos Grupos Flujo de Actividad Relacionados

Título

Descripción

Conjunto de datos de Ejemplo
Este conjunto es solo un ejemplo

datos ejemplo

Información Adicional

Campo	Valor
Fuente	http://www.interior.gob.es/
Autor	Ministerio del Interior
Estado	active
Frecuencia actualización	anual

Para conocer más posibilidades sobre la gestión de conjuntos de datos u organizaciones se puede consultar la [guía de usuario](#) de CKAN.

EVALUACIÓN DE USO

Una vez hemos publicado nuestros datos es interesante también tener cierta **medida del uso** que se les está dando para saber cuáles pueden ser los más interesantes desde el punto de vista de quienes los van a reutilizar. Esto nos ayudará también a definir mejor nuestros criterios de selección y priorización en la

publicación de datos. El propio CKAN nos proporciona algunas herramientas que pueden también ser de utilidad para este objetivo, como:

- La [extensión de estadísticas](#) del catálogo, accesible añadiendo `/stats` a la dirección de nuestro catálogo, que nos permitirá conocer entre otros:
 - o El número de conjuntos de datos publicados.
 - o El número de actualizaciones semanales.
 - o Los conjuntos de datos mejor valorados.
 - o Los conjuntos de datos con más actualizaciones.



- La [extensión de Google Analytics](#) para CKAN, que nos ayudará a integrar ambos sistemas y contar con distintos indicadores como *número de visitas*, *descargas*, *acceso a la API*, *distintas analíticas de datos*, etc.
- El [monitor de visitas](#) a páginas que, una vez activado, nos ayudará a:
 - o Ordenar los conjuntos de datos por popularidad.
 - o Destacar conjuntos de datos y recursos populares.
 - o Mostrar contadores de visitas para conjuntos de datos y recursos.
 - o Exportar los datos para su análisis.

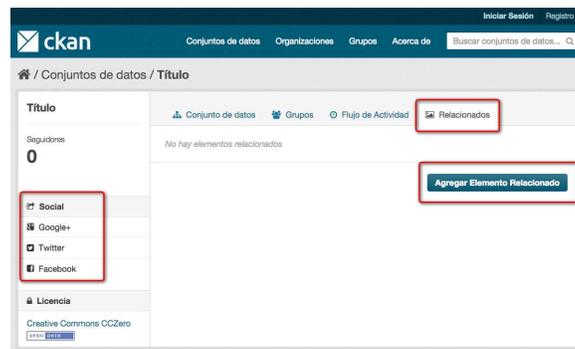
DIÁLOGO CON LA COMUNIDAD

Además de los indicadores de evaluación exclusivamente cuantitativos que acabamos de ver, no debemos olvidar que para ser capaces de comprender la demanda de datos en profundidad deberíamos también establecer un diálogo continuo con la **comunidad de reutilizadores** de los datos, entendiendo como tal al conjunto de:

- Administraciones públicas a cualquier nivel.
- Organizaciones civiles, activistas y ciudadanos en general.
- Sector académico y de la innovación.
- Infomediarios* y empresas.
- Profesionales de los datos y especialistas.
- Medios de comunicación.

Nuevamente CKAN nos ofrece algunas herramientas interesantes para ayudarnos con el diálogo y la comunicación, entre las que podemos destacar:

- Las funciones que proporciona para compartir conjuntos de datos en la redes sociales más populares (*Google+*, *Twitter* y *Facebook*).
- La posibilidad de comentar y discutir sobre los conjuntos de datos gracias a extensiones como [disqus](#)⁹ o [comments](#).
- La opción de [compartir ideas](#), aplicaciones o cualquier otro contenido relacionado con los datos publicados.



Sin embargo, conviene también recordar que, para poder llevar a cabo los objetivos que se planteen, hará falta combinar estas herramientas *online* con otras actuaciones *offline* necesarias para **facilitar la comunicación y colaboración continua**, así como para **conocer de primera mano sus necesidades** y obtener *feedback* útil a través de otros **indicadores cualitativos** como por ejemplo *cuestionarios, entrevistas, grupos de trabajo, etc.*

MANTENIMIENTO

La carga inicial de los datos no es más que el primer paso en la apertura de la información. La incorporación de nuevos conjuntos de datos debería ser una constante a partir de ese momento, y especialmente la actualización frecuente de los conjuntos de datos que ya hayan sido incorporados a la plataforma. Para ello tendremos en cuenta que contamos con dos tipos de conjuntos de datos:

- **Estáticos**: Cuando los datos están disponibles a través de **archivos** en cualquier formato. Esta opción requerirá que alguien se encargue de su mantenimiento actualizando el conjunto de datos de forma manual *subiendo* un nuevo archivo de forma periódica con la frecuencia que se haya establecido en cada caso.
- **Dinámicos**: Cuando se pueda establecer una **conexión directa** con la fuente original desde donde se obtienen los datos a través de una dirección web (URL) donde se han publicado. En este caso cada vez que los datos se

⁹ A través del servicio externo proporcionado por <https://disqus.com/>



actualicen se reflejará directamente en la plataforma de catalogación sin necesidad de intervención alguna.

EVOLUCIÓN

Para finalizar, debe quedar claro que el objetivo de esta guía es principalmente ayudar a todos aquellos que cuenten con pocos recursos y conocimientos limitados a dar sus primeros pasos y experimentar con la publicación de datos abiertos. Las indicaciones aquí recogidas servirán para poder experimentar fácilmente en muy poco tiempo y sin apenas necesidad de inversión.

No obstante, si se quiere apostar por una estrategia continuada de apertura de datos a medio y largo plazo, será necesario pensar en llevar a cabo un proyecto más complejo prestando especial atención a:

- La elaboración de una [estrategia y planificación](#) global que nos ayude a establecer claramente nuestros objetivos y a llevarlos a cabo.
- El aprovisionamiento de los recursos (humanos y económicos) necesarios para poder llevarla a cabo.
- El cumplimiento estricto de la [NTI de reutilización de recursos de información](#) y cualquier otra normativa aplicable. Algo en lo que la [Guía de Aplicación del Real Decreto 1495/2011](#) puede resultar muy útil.
- Un adecuado dimensionamiento de la tecnología (software y hardware) empleada, incluyendo cuestiones de seguridad y escalabilidad.
- La importancia de participar con y de la Comunidad Open Data y en proyectos vertebradores como la iniciativa **Aporta y el [Catálogo Nacional](#)**.

REFERENCIAS

Aragón Open Data, Guía de catalogación de datos abiertos.

http://opendata.aragon.es/public/documentos/Guia_Gestion_v1.0.pdf

EU, Measuring European Public Sector Information Resources.

http://ec.europa.eu/information_society/newsroom/cf/document.cfm?doc_id=1198

Datos.gob.es, Guía de aplicación del Real Decreto 1495/2011.

<http://datos.gob.es/sites/default/files/PLANCISP-GRD-07.3.41.pdf>

Datos.gob.es, Guía de catalogación de conjuntos de datos.

http://datos.gob.es/sites/default/files/sites/default/files/Guia_catalogacion_datosgobes_v3.pdf

ONTSI, Estudio de caracterización del sector infomediario en España.

<http://www.ontsi.red.es/ontsi/es/estudios-informes/estudio-de-caracterizaci%C3%B3n-del-sector-infomediario-en-espa%C3%B1a-edici%C3%B3n-2012>

UK Cabinet Office, G8 Open Data Charter and Technical Annex.

<https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex>

EC, Guidelines on recommended standard licenses, datasets and charging.

http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?action=display&doc_id=6421

Open Data Barometer, OGD Implementation: data availability.

<http://opendatabarometer.org/report/analysis/implementation.html>

Datos.gob.es, Taxonomía del Catálogo.

http://datos.gob.es/datos/sites/default/files/files/12_tax_02.pdf

BOE, Real Decreto 1495/2011, por el que se desarrolla la Ley 37/2007 sobre reutilización de la información del sector público.

http://www.boe.es/diario_boe/txt.php?id=BOE-A-2011-17560

Open Definition, Conformance licenses.

<http://opendefinition.org/licenses/>

Creative Commons, About the Licenses.

<http://creativecommons.org/licenses/>

LAPSI, License interoperability report.

http://lapsi-project.eu/sites/lapsi-project.eu/files/D5_1_Licence_interoperability_Report_final.pdf

LAPSI, Licensing guidelines.

[http://www.lapsi-project.eu/sites/lapsi-project.eu/files/D5.2LicensingGuidelinesPO%20\(1\).pdf](http://www.lapsi-project.eu/sites/lapsi-project.eu/files/D5.2LicensingGuidelinesPO%20(1).pdf)

Datos.gob.es, Normativa.

<http://datos.gob.es/saber-mas?q=taxonomy/term/28>

Datos.gob.es, Buenas prácticas de formatos para la publicación de conjuntos de datos.

<http://datos.gob.es/sites/default/files/files/SoporteRISP-Buenas%20pr%C3%A1cticas%20formatos%20de%20publicaci%C3%B3n.docx>

Datos.gob.es, Guía metodológica para iniciativas de datos abiertos sectoriales.

<http://datos.gob.es/content/saber-mas/guia-metodologica-iniciativas-de-datos-abiertos-sectoriales>

Datos.gob.es, Informe de Tendencias e Iniciativas de Datos Abiertos.

<http://datos.gob.es/content/informe-de-tendencias-iniciativas-de-datos-abiertos>

Amazon, Amazon Web Services.

<http://aws.amazon.com/>

CKAN, User guide.

<http://docs.ckan.org/en/latest/user-guide.html>

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>

Datos.gob.es, Guía de aplicación de la NTI de Reutilización de recursos de información.

<http://datos.gob.es/content/guia-de-aplicacion-de-norma-tecnica-de-interoperabilidad-de-reutilizacion-de-recursos-de>

Aragón Open Data, Informe de conformidad con la NTI de Reutilización de recursos de información.

http://opendata.aragon.es/public/documentos/Informe_NT_I_Aragon_OpenData_v31-01-14.pdf

Wolters Kluwer, Monográfico especial Gobierno Abierto y Open Data

https://funkziuni.files.wordpress.com/2014/07/aaj_7-8_2014_br.pdf