

MANUAL DATOS ABIERTOS EN COLOMBIA

PARA:

FEDERACIÓN

En el siguiente documento se explica el procedimiento para realizar las diferentes acciones que posibilita la herramienta Socrata



MINTIC

ASESOFTWARE
25 AÑOS



vive digital
para la gente



**TODOS POR UN
NUEVO PAÍS**
PAZ EQUIDAD EDUCACIÓN

TABLA DE CONTENIDOS

1. PROPÓSITO DEL DOCUMENTO.....	1
2. ¿QUÉ ES FEDERACIÓN DE DATOS ABIERTOS?	2
3. ¿CÓMO FUNCIONA LA FEDERACIÓN?.....	3
3.1. BASADO EN SOCRATA.....	3
3.2. BASADO EN CKAN.....	7
3.2.1. INSTALACIÓN.....	8
3.2.2. GENERAR /DATA.JSON OFF-LINE	9
3.2.3. SOCRATA - CKAN	10
4. ¿QUÉ BENEFICIOS TIENE FEDERAR?	11
5. PREGUNTAS FRECUENTES.....	12
5.1. MODERANDO UNA VISTA	12
5.2. IMPORTANCIA DE UN CATÁLOGO FEDERADO.....	13
5.3. ENRUTAMIENTO Y AUTORIZACIÓN EN FEDERACIÓN.	13



01

PROPÓSITO DEL DOCUMENTO

En el presente documento se pretende recoger información de utilidad, a modo de tutorial, para los usuarios administradores del portal de datos abiertos del estado colombiano (Socrata) acerca del proceso de federación.

La federación permite compartir conjuntos de datos con los ciudadanos mediante la agregación de catálogos de datos de otros portales diseñados en Socrata o en proyectos basados en CKAN, los cuales se pueden integrar al portal de datos abiertos del estado colombiano.



02

¿QUÉ ES FEDERACIÓN DE DATOS ABIERTOS?

Los Datos Abiertos tienen éxito cuando la información es compartida y distribuida a la mayor cantidad de ciudadanos posible. Lo anterior tiene sentido, dado que los datos se originan de múltiples entidades nacionales y municipales a varios niveles en todo el país, e evidencia una necesidad de desarrollar una solución para propagar datos a través de varias entidades u organizaciones sin duplicarlos ni incrementar el almacenamiento requerido.

Socrata ofrece *Open Data Federation Services (ODFS)*, descrita como una arquitectura y un conjunto de herramientas automatizadas, que proporcionan de manera sencilla y transparente que las entidades puedan compartir datos en sus propios catálogos provenientes de otros catálogos de datos abiertos. *Socrata Open Data Federation Services* puede ser configurados de varias maneras:

- Un dominio de origen puede exponer una parte o la totalidad de su catálogo de datos abiertos a otro dominio destino.
- Un dominio puede actuar exclusivamente como una fuente para exponer conjuntos de datos, exclusivamente como un destino de exposición, o ambos, como una fuente y un destino.

Socrata también proporciona integración con otras herramientas de gestión de catálogos como CKAN.

03

¿CÓMO FUNCIONA LA FEDERACIÓN?

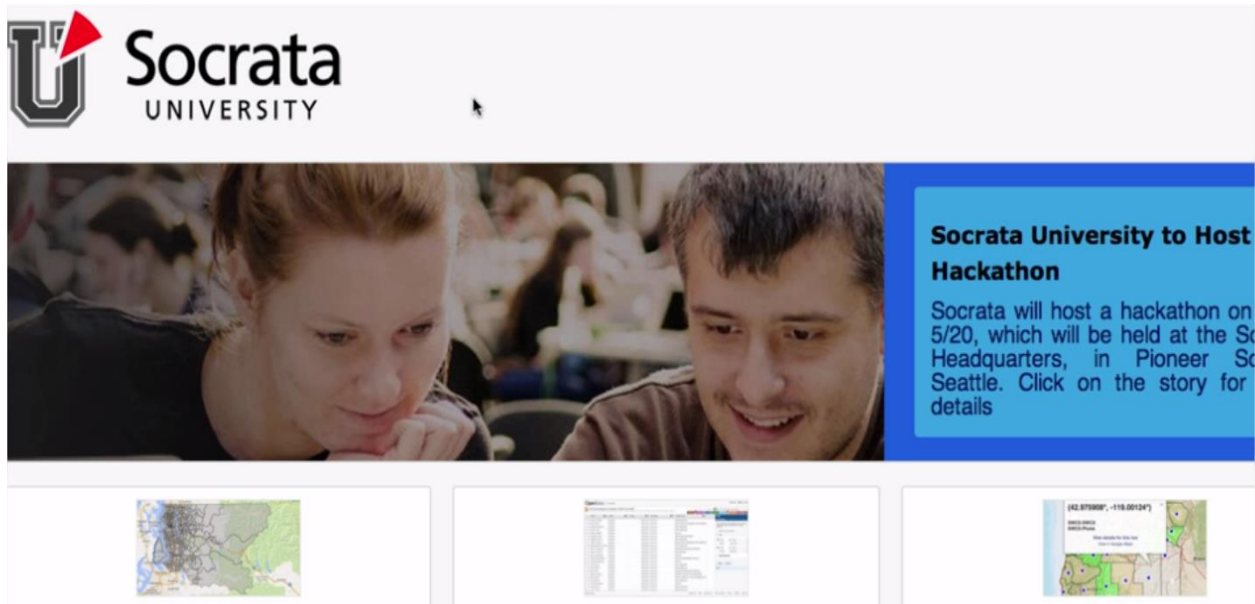
A continuación, presentaremos la federación tanto entre sitios web basados en Socrata, como con la herramienta de gestión de catálogos de datos CKAN.

3.1. BASADO EN SOCRATA

Un dominio de socrata federa otros dominios de socrata. Y se presenta en su catálogo en el área de Dominios Federados.

The screenshot shows a Socrata catalog interface. On the left, there are two dropdown menus: 'Topics' and 'Federated Domains'. The 'Federated Domains' menu is highlighted with a blue border and contains the following options: 'This site only', 'data.baltimorecity.gov', 'data.montgomerycoun...', 'opendata.howardcoun...', and 'queenannescounty.dat...'. The main content area displays three data records under the 'Public Safety' category, all provided by 'opendata.howardcountymd.gov'. The records are: 1. 'Fire & Rescue Incidents: 2012 - Present' (Updated September 28, 2015, 1,538,007 views). 2. 'Fire & Rescue Personnel Responses: 2012 - 2014' (Updated May 01, 2015, 1,527,298 views). 3. 'Fire Dispatches: 2012 - 2014' (Updated May 01, 2015). Each record includes a description, a 'More' link, and 'API Docs'.

Un usuario administrador de un dominio debe crear una solicitud, para esto veremos un ejemplo con el portal de *Socrata University*.



El usuario administrador de la plataforma debe iniciar sesión.

Sign In to Socrata

 Sign in with your **Socrata ID**

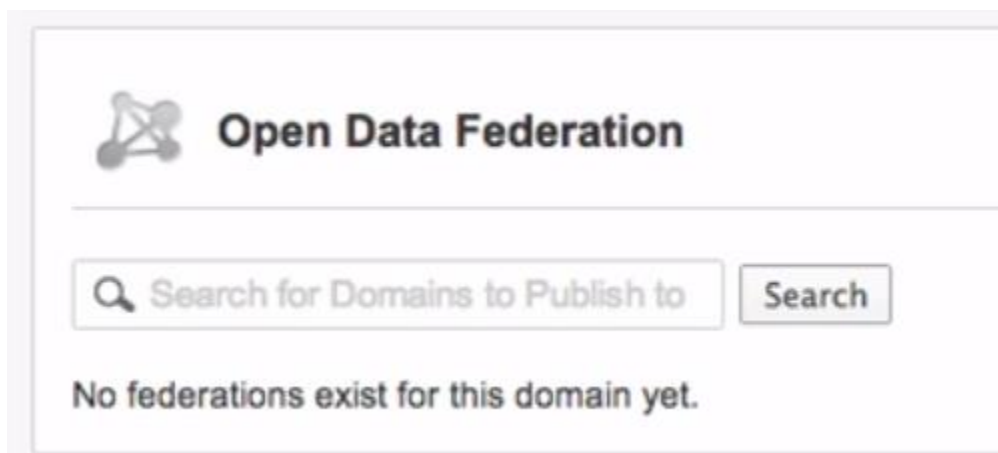
Use your Email and Password to sign into all **Socrata powered** sites.

Email Address

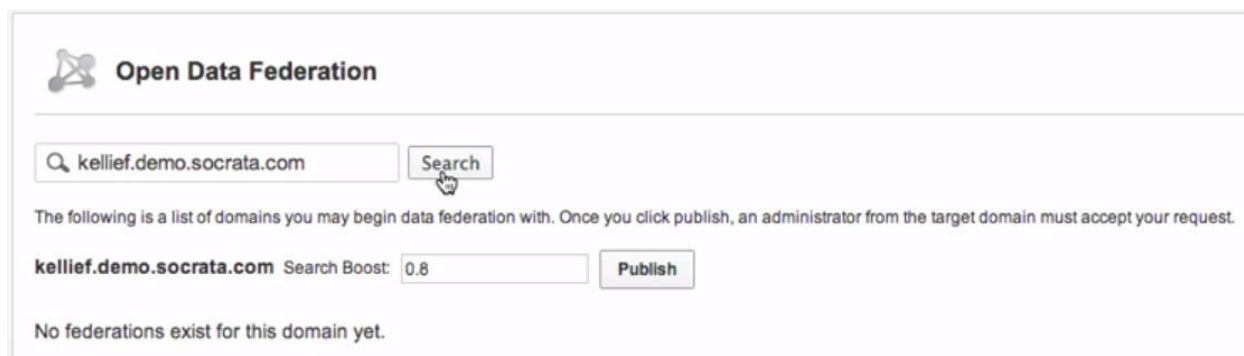
Password [Forgot Password?](#)

[Sign In](#) Don't have an account? [Sign up now.](#)

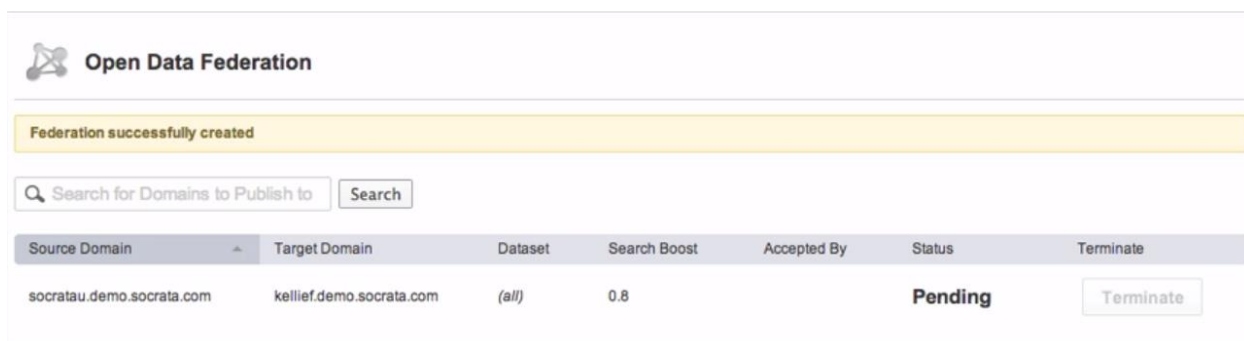
Ingresar al panel de *Administración* y luego *Federación*.



Especificar en el campo, el dominio Socrata al cual se le va a hacer la petición de federación, colocarlo en el espacio indicado y luego botón *Buscar*.

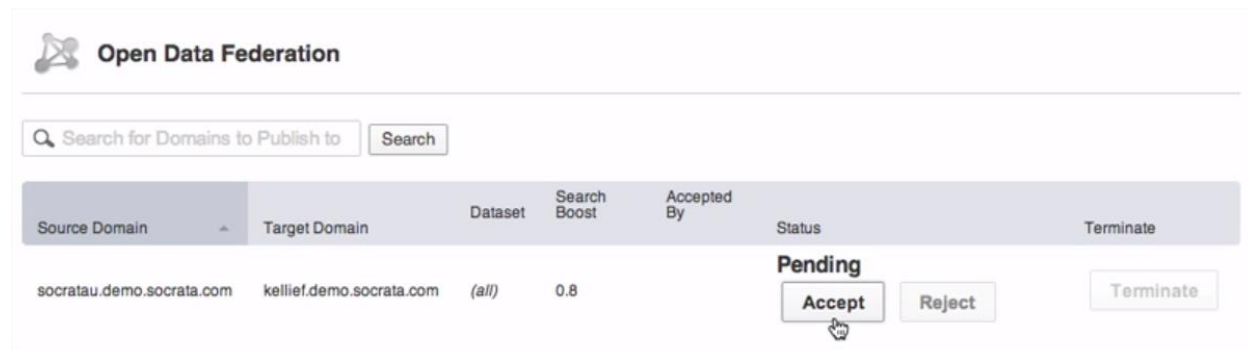


Socrata mostrara si el dominio está correcto, posteriormente se puede lanzar la petición por medio del botón *Publicar*.



Socrata enviara la petición de federación al dominio destino, y mostrara que el estatus de la misma se encuentra en *Pendiente*, en cualquier momento el dominio fuente puede terminar esta acción de federar, por medio del botón *Terminar*.

El administrador del portal de destino de la federación, una vez iniciado sesión en su portal, ingresando por el panel de *Administración* y posteriormente *Federación*, podrá ver la petición realizada por el dominio fuente, y será tarea del administrador aceptar o rechazar esta petición. Clic en *Aceptar*.



The screenshot shows the 'Open Data Federation' interface. At the top, there is a search bar with the text 'Search for Domains to Publish to' and a 'Search' button. Below this is a table with the following columns: Source Domain, Target Domain, Dataset, Search Boost, Accepted By, Status, and Terminate. A single row is visible with the following data: Source Domain: socratau.demo.socrata.com, Target Domain: kellief.demo.socrata.com, Dataset: (all), Search Boost: 0.8, Status: Pending. Under the 'Status' column, there are three buttons: 'Accept', 'Reject', and 'Terminate'. A mouse cursor is pointing at the 'Accept' button.

Después de aceptarla, se podrá ver en el área de *Dominios Federados* que se ha adicionado la nueva petición de federación del dominio Socrata University. Al ingresar a esta federación veremos la lista de conjuntos federados y la fuente de los mismos.



The screenshot shows a search results page for federated domains. At the top, it says '126 Results filtered by Federated Domains > data.baltimorecity.gov x' and 'Sort by Most Accessed'. On the left, there is a sidebar with 'Categories' and 'View Types'. The main content area shows two dataset cards. The first card is titled 'BPD Part 1 Victim Based Crime Data' and is categorized under 'Public Safety'. It includes a description: 'All BPD data on Open Baltimore is preliminary data and subject to change. The information presented through Open Baltimore represents Part I victim based...' and a 'More' link. It also shows 'Updated August 25, 2016' and 'Views 42,246'. The second card is titled 'Baltimore City Employee Salaries FY2015' and is categorized under 'City Government'. It includes a description: 'This database captures gross salary from July 1, 2014 through June 30, 2015 and includes only those employees who were employed on June 30, 2015' and shows 'Updated April 20, 2016' and 'Views'.

Si queremos ver el detalle del conjunto al darle clic sobre el título del mismo, inmediatamente se hace la redirección hacia la fuente.

OPEN BALTIMORE

BPD Part 1 Victim Based Crime Data
All BPD data on Open Baltimore is preliminary data and subject to change. The information presented through Open Baltimore ▶

	CrimeDate	CrimeTime	CrimeCode	Location	Description	Inside/Outside	Weapon	Post	District	Neighborhood
1	08/20/2016	19:34:00	4E	900 N CALHOUN ST	COMMON ASSAULT	O	HANDS	743	WESTERN	Sandl
2	08/20/2016	19:42:00	5A	100 S SCHROEDER ST	BURGLARY	I		931	SOUTHERN	Hollin
3	08/20/2016	20:16:00	5D	1600 SPENCE ST	BURGLARY	O		831	SOUTHWESTERN	Morre
4	08/20/2016	04:30:00	4D	700 N ELLWOOD AVE	AGG. ASSAULT	O	HANDS	323	EASTERN	Madis
5	08/20/2016	00:30:00	4E		COMMON ASSAULT	I	HANDS	321	EASTERN	
6	08/20/2016	04:00:00	4C	1400 BLOOMFIELD AVE	AGG. ASSAULT	I	OTHER	832	SOUTHWESTERN	Violet
7	08/20/2016	15:00:00	5A	800 N FULTON AVE	BURGLARY	I		722	WESTERN	Harle
8	08/20/2016	15:55:00	6D	4100 GROVELAND AVE	LARCENY FROM AUTO	O		621	NORTHWESTERN	West.

3.2. BASADO EN CKAN

Es una herramienta de uso simple, que sincroniza periódicamente un catálogo de CKAN en una instancia de Socrata. Un ejemplo de lo anterior lo podemos encontrar en: <https://www.data.gov/issue/> este sitio realizado por Socrata, funciona con catálogos basados en proyectos CKAN.

Desde el proyecto de CKAN, usando una extensión e genera un /data.json (u otra ruta configurable) que da salida a los contenidos de un catálogo de datos, en formato Open Data JSON.

Este tutorial asume metadatos que se almacenan en CKAN como lo hacen: <http://hub.healthdata.gov/>, si usted está almacenando metadatos de manera diferente, tendrá que revisar ckanext /datajson /plugin.py .

Para más información sobre el proceso se deja el siguiente link <https://github.com/HHS/ckanext-datajson>.

3.2.1.INSTALACIÓN

Activar CKAN virtualenv, instalar dependencias e instalar el módulo en modo desarrollador, con lo anterior se debe colocar en la ruta Python.

```
. path/to/pyenv/bin/activate  
pip install -r pip-requirements.txt  
python setup.py develop
```

En el archivo CKAN.ini, se debe adicionar “datajson” con la línea ckan.plugins:

```
ckan.plugins = (other plugins here...) datajson
```

Este es el plug-in para la salida /data.json. Para realizar la federación, se debe adicionar:

```
ckan.plugins = (other plugins here...) harvest datajson_harvest
```

Si se ejecuta CKAN a través de WSGI, se puede encontrar un error de dependencia. La solución sería adicionar:

```
import ckanext
```

Antes de:

```
from paste.deploy import loadapp
```

A continuación, reinicie el servidor y este muy pendiente de:

```
http://yourdomain.com/data.json  
and  
http://yourdomain.com/data.jsonld  
and  
http://yourdomain.com/pod/validate
```

3.2.2.GENERAR /DATA.JSON OFF-LINE

La generación de este archivo es un poco lenta, por lo que una alternativa en lugar de almacenamiento en caché es generar el archivo periódicamente (por ejemplo, en una tarea programada). En ese caso, tendrá que cambiar la ruta que genera el archivo CKAN de otro modo. En el archivo CKAN.ini, en la aplicación sección principal, añadir:

```
ckanext.datajson.path = /internal/data.json
```

Ahora cree un archivo crontab ("mycrontab") para descargar este URL a un archivo en el disco cada diez minutos:

```
0-59/10 * * * * wget -qO /path/to/static/data.json http://localhost/internal/data.json
```

Y activar tu crontab:

```
crontab mycrontab
```

En Apache, debemos bloquear el acceso desde el exterior a la URL "interno". En su httpd.conf, añada:

```
Alias /data.json /path/to/static/data.json

<Location /internal/>
    Order deny,allow
    Allow from 127.0.0.1
    Deny from all
</Location>
```

Reinicie Apache. Esperar a que el trabajo de cron se ejecute una vez, a continuación, se debe comprobar las cargas /data.json (y debe ser rápido).

3.2.3.SOCRATA - CKAN

Se debe generar un “*recolector*” para permitir que los directorios de CKAN puedan mantenerse sincronizados con Socrata.

Es necesarios construir y usar este plugin:

```
git clone git@github.com:socrata/socrata-harvester.git
cd socrata-harvester
pip install -r pip-requirements.txt
python setup.py develop
```

Actualice CKAN para incluir este “*recolector*” adicionando el plug-in *socrata_harvester*.

```
ckan.plugins = harvest socrata_harvester
```

Después de configurar todo lo anterior, usted debería ser capaz de ir a: <http://localhost:5000/recolector> y tiene un nuevo tipo de federación. Para más información dirigirse al siguiente link <https://github.com/socrata/socrata-harvester#using>



04

¿QUÉ BENEFICIOS TIENE FEDERAR?

- Sustituye procesos manuales y elimina duplicidad de esfuerzos de las entidades publicadoras y sus sitios de datos abiertos.
- Se asegura la conservación, sin ninguna intervención manual.
- Amplia el alcance de www.datos.gov.co a un alto nivel de reconocimiento de marca.
- Los ciudadanos pueden localizar los datos y satisfacer sus necesidades de información, basados en temas y preguntas que trascienden las fronteras de una entidad publicadora.
- Los desarrolladores tienen una forma programática coherente para encontrar y recuperar los datos sin tener que lidiar con múltiples fuentes de datos, desconectados.

5.1. MODERANDO UNA VISTA

Ambos dominios permiten moderar una vista.

Si los dos dominios permiten moderar la vista, entonces todos los puntos de vista que han sido aprobados desde el dominio de origen aparecerán en el catálogo, y no debe ir a través de la moderación vista en el dominio de destino.

El dominio de origen ha habilitado la moderación vista, pero el dominio de destino no lo hace.

Cualquier punto de vista aprobados desde el dominio de origen aparecerán en el dominio de destino, y no deberá ir a través de un punto de moderación vista adicional, ya que no está habilitado en el dominio de destino.

El dominio de origen no tiene habilitado la moderación vista, pero el dominio de destino sí.

Ninguna de las vistas desde el dominio de origen aparecerá en el dominio de destino, y no se enumeran en vista de la moderación para su aprobación.

Ni la fuente ni el dominio de destino han permitido a la moderación vista.

Todos los puntos de vista del dominio de origen aparecerán en el dominio de destino, y no van a ir a través de la moderación vista.

5.2. IMPORTANCIA DE UN CATÁLOGO FEDERADO

Cuando se realiza la federación de un catálogo de entidades pequeñas (como la federación de un catálogo estado con el catálogo de ciudades y / o condados), es posible que desee ajustar la prioridad de los conjuntos de datos que muestran en las consultas de búsqueda.

Puede configurar la federación en una opción de "Buscar Boost" que determina qué tan alto aparece en las búsquedas. El sitio principal aparecerá siempre en primer lugar, y luego el "Buscar Boost" es una puntuación entre 0-1, con valores cercanos a 1 que aparece anteriormente.

No se puede cambiar el impulso de búsqueda una vez configurada para cada federación.

5.3. ENRUTAMIENTO Y AUTORIZACIÓN EN FEDERACIÓN.

El dominio de destino tiene enrutamiento y aprobación establecidos.

Para los dominios con R&A establecidos, sólo los conjuntos de datos que han sido aprobados desde el dominio de destino, junto con sus puntos de vista derivados, aparecerán en el catálogo. Vistas derivadas de conjuntos de datos no aprobados o rechazados no serán visibles. Conjuntos de datos y sus puntos de vista y derivados del dominio de origen tienen que pasar por el proceso de enrutamiento y la aprobación del objetivo para ser visible en el catálogo.

El dominio de destino no tiene la aprobación y enrutamiento establecidos, pero la fuente sí.

Todos los conjuntos de datos y vistas derivadas del dominio de destino serán visibles en el catálogo. Conjuntos de datos y vistas derivadas del dominio de origen sólo será visible en el catálogo si el dominio de origen ha aprobado el conjunto de datos.

Ni el dominio de destino, ni el dominio de origen tienen aprobación y enrutamiento establecidos para este compromiso.

Todos los conjuntos de datos y vistas derivadas tanto del destino y de origen dominios serán visibles en el catálogo.