

MANUAL DEL USUARIO PUBLICADOR



El portal centralizado de datos del Estado

Versión 01 | Diciembre 2020

CONTENIDOS

¿Qué es datos.gob.cl?	3
¿Para qué sirve datos.gob.cl?	3
¿Cómo puede participar mi institución?	3
¿Quién es el responsable de cargar los datos de mi institución?	4
¿Qué aspectos legales deben considerar las instituciones?	4
¿De qué otras tareas se debe encargar mi institución para cargar datos en datos.gob.cl?	4
¿Qué es un formato abierto?	4
¿Cómo puedo consumir datos de las instituciones disponibles en datos.gob.cl?	5
¿Cuáles son mis responsabilidades como editor?	5
¿Cómo habilito mi institución en datos.gob.cl	5
Uso del portal datos.gob.cl	6
Login en el sistema	6
Mi perfil	7
Agregar un conjunto de datos (Proceso manual)	8
Editar un conjunto de datos	10
Editar un Recurso	12
Eliminar un Recurso	13
Proceso de publicación utilizando API	14
Proceso de publicación usando Pentaho en el portal de datos	21

¿Qué es datos.gob.cl?

Datos.gob.cl es el repositorio de datos abiertos del Estado, en el cual las instituciones publican sus datos de manera sencilla y transparente, en formatos abiertos, bajo la lógica de un catálogo centralizado y ordenado, de rápida búsqueda y fácil uso.

¿Para qué sirve datos.gob.cl?

El portal datos.gob.cl al ser un repositorio centralizado de datos de las instituciones públicas, sirve para poner a disposición de diferentes actores datos generados con presupuesto público y en formatos abiertos, para libre consumo de los interesados. Estos datos se pueden consumir descargando desde la plataforma o consumiendo directamente desde la API que genera datos.gob.cl. Permitiendo la generación de valor como visualizaciones, aplicaciones, estudios, investigaciones y servicios, entre otros.

¿Cómo puede participar mi institución?

- Revisar y aceptar los términos y condiciones del portal datos.gob.cl disponibles en datos.gob.cl/terminos-y-condiciones.
- Solicitar cuenta de usuario a la División de Gobierno Digital: Para obtener una cuenta de usuario en el sitio datos.gob.cl, las instituciones deben realizar la solicitud de cuenta en el siguiente enlace: [solicitud de cuenta](#). Este link también se puede encontrar en el portal datos.gob.cl, en la sección de "Ayuda institucional".
- Un la solicitud, la institución solicitante deberá ingresar con su Claveúnica y completar los siguientes datos:
 - Nombre institución
 - Correo electrónico institucional
 - Aceptar los términos y condiciones
- La División de Gobierno Digital, realizará una validación del solicitante con el coordinador de Transformación Digital. En este paso, la solicitud podrá ser aceptada o rechazada.
- Una vez realizada la validación de los datos enviados por parte de la División de Gobierno Digital, en caso de ser aceptada, se enviará una invitación al usuario para sumarse al portal datos.gob.cl. En caso de ser rechazada, se enviará un correo indicando los motivos por los cuales fue rechazada.
- Con el nombre de usuario asignado, deberá crear una contraseña para poder acceder al portal.

¿Quién es el responsable de cargar los datos de mi institución?

El responsable de la carga de datos puede ser designado por diferentes áreas, por ejemplo puede ser alguien de transparencia, tecnología, comunicaciones, estudios, etc.

¿Qué aspectos legales deben considerar las instituciones?

Considerando que las instituciones públicas manejan datos personales, es importante que la publicación que realicen en el portal datos.gob.cl, estén alineados con lo propuesto por la ley 19.628 sobre protección de la vida privada, evitando la publicación de información personal o sensible.

Es importante que las instituciones públicas puedan asignar una licencia que permita el uso y la reutilización sin restricciones por parte de los actores interesados. Esto, debido a que la ley 17.336 por el solo hecho de la creación de una obra, protege los derechos de los autores de la misma y sus derechos conexos.

Por otro lado, se debe considerar lo planteado por la ley 20.285 sobre acceso a la información pública, que indica, cuáles son los criterios para no hacer pública información en posesión de las instituciones. Los datos e información pública que no estén dentro de estas causales pueden ser publicados en formatos abiertos para su reutilización.

¿De qué otras tareas se debe encargar mi institución para cargar datos en datos.gob.cl?

La institución previo a la carga de datos, se debe encargar de anonimizar las bases de datos que cuenten con información personal o sensible. Además, debe preocuparse por publicar la data lo más desagregada posible, con el menor grado de procesamiento, evitando poner imágenes o gráficos. Debe encargarse de dejar a disposición la información adicional que pueda servir de contexto, ya sea, informes o diccionarios de variables.

Por último, debe facilitar un correo de contacto del encargado institucional de la publicación de datos en el portal, al cual poder hacer llegar las preguntas de los usuarios relacionadas a los conjuntos de datos publicados. Este correo debe quedar señalado en los metadatos de cada conjunto de datos publicados.

¿Qué es un formato abierto?

Según lo señalado por el Open Data Handbook, se entiende formato abierto como: "aque

donde las especificaciones del software están disponibles para cualquier persona, de forma gratuita, así cualquiera puede usar dichas especificaciones en su propio software sin ninguna limitación en su reutilización que fuere impuesta por derechos de propiedad intelectual” (Open Data Handbook).

¿Cómo puedo consumir datos de las instituciones disponibles en datos.gob.cl?

Para consumir datos, puedes acceder sin usuarios ni contraseñas, solo debes ingresar a datos.gob.cl y descargar los conjuntos de datos o consumir directamente su API.

¿Cuáles son mis responsabilidades como editor?

Como editor, debo asegurarme que los datos publicados posean la calidad y pertinencia necesaria para su reutilización. Además, de publicar los datos con la periodicidad comprometida por la institución. Por último, es necesario definir una licencia adecuada para el uso por parte de los interesados.

¿Cómo habilito mi institución en datos.gob.cl

Datos.gob.cl es el repositorio de datos abiertos centralizado del Estado, en el cual las instituciones pueden publicar sus datos en diversos formatos, bajo la lógica de un catálogo, con rápida búsqueda y permitiendo la reutilización de los conjuntos publicados. Permitiendo la generación de valor mediante la creación de visualizaciones, aplicaciones, estudios, investigaciones, servicios, entre otros.

Las instituciones pueden utilizar el portal datos.gob.cl o trabajar en un portal propio de publicación de datos. Sin embargo, si desean sumarse a la iniciativa centralizada, deben realizar los siguientes pasos:

1. Revisar los términos y condiciones del portal datos.gob.cl, estos los pueden encontrar en el siguiente enlace: [LINK DE TÉRMINOS Y CONDICIONES](#)
2. Solicitar cuenta de usuario a la División de Gobierno Digital: Para obtener una cuenta de usuario en el sitio datos.gob.cl, las instituciones deben realizar la solicitud de cuenta en el siguiente enlace: [solicitud de cuenta](#). Este link también se puede encontrar en el portal datos.gob.cl, en la sección de “Ayuda institucional”.
3. En la solicitud, la institución solicitante deberá ingresar con su Claveúnica y completar los siguientes datos:
 - Nombre institución
 - Correo electrónico institucional
 - Aceptar los términos y condiciones

4. La División de Gobierno Digital, realizará una validación del solicitante con el coordinador de Transformación Digital. En este paso, la solicitud podrá ser aceptada o rechazada.
5. Una vez realizada la validación de los datos enviados por parte de la División de Gobierno Digital, en caso de ser aceptada, se enviará una invitación al usuario para sumarse al portal datos.gob.cl. En caso de ser rechazada, se enviará un correo indicando los motivos por los cuales fue rechazada.
6. Con el nombre de usuario asignado, deberá crear una contraseña para poder acceder al portal.

Uso del portal datos.gob.cl

Para poder acceder a las funcionalidades del portal datos.gob.cl es necesario contar con un usuario en el sistema y encontrarse logueados. Las vistas de administración del sistema cuenta un diseño diferente a las vistas públicas la principal diferencia es que en estas el navbar y footer tienen un diseño distinto.

Login en el sistema

Para hacer login en el sistema es necesario ir a “Iniciar sesión” en la parte superior derecha.



De esta forma se accede a la siguiente vista donde el usuario puede ingresar su usuario y contraseña para hacer login en el sistema.

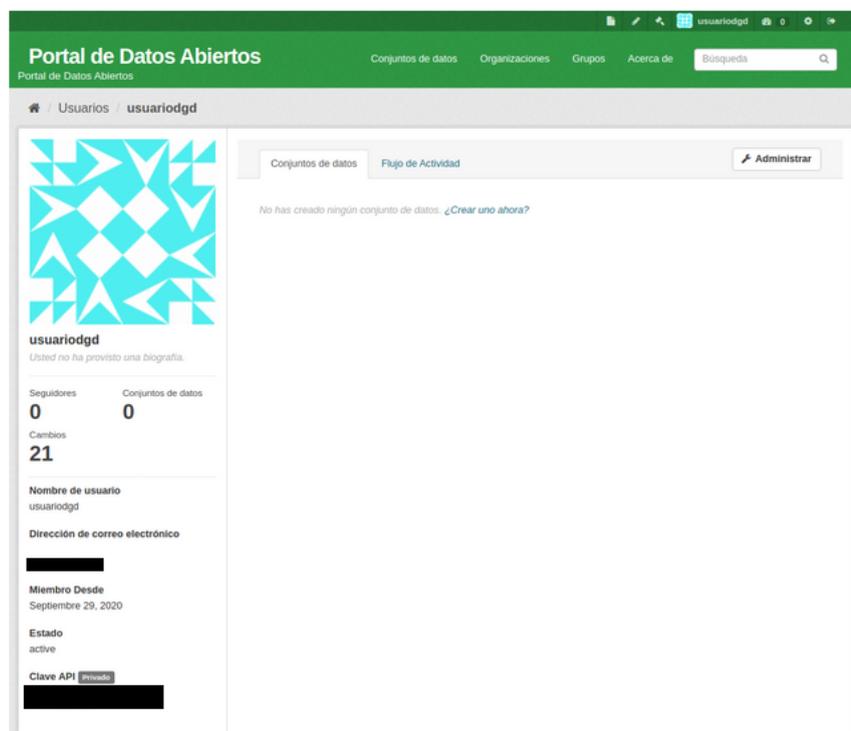
The image shows the login page of datos.gob.cl. At the top is the same navigation bar as in the previous screenshot. Below it is the heading 'Conectarse'. There are two input fields: 'Nombre de usuario' with the placeholder 'Enter your email' and 'Contraseña' with the placeholder 'Enter your password'. Below the password field is a checked checkbox labeled 'Recordarme'. To the right of the input fields is a link '¿Olvidaste tu contraseña?' with the text 'No hay problema, utiliza nuestro formulario de recuperación de contraseña para restablecer.' and a button 'Restablecer Contraseña'. At the bottom left is a blue button labeled 'Conectarse'.

Una vez que el usuario haya iniciado sesión, en la barra de navegación aparecerá su nombre de usuario y al hacer clic en él se desplegará un menú para acceder a las distintas funcionalidades con las que cuenta el usuario.



Mi perfil

En esta vista aparece la información del usuario ingresado, al costado derecho aparece el nombre de usuario junto con el avatar de este usuario, bajo este recuadro aparece información del sitio asociada al usuario. En la parte inferior de este recuadro en la sección "Clave API" aparece el token para que este usuario pueda ocupar todas las funcionalidades de la API de CKAN.



En la pestaña de Conjunto de datos se muestran los distintos conjuntos de datos creados por el usuario en el sistema y en la vista de "Flujo de Actividad" se muestran las distintas acciones que ha realizado el usuario en el sistema. Si se selecciona la opción de "Administrar" se redirige a la siguiente vista:

En esta vista el usuario puede modificar su información, el administrador del sistema puede actualizar la contraseña y regenerar la Clave API del usuario.

Agregar un conjunto de datos (Proceso manual)

Un conjunto de datos puede ser agregado de más de una forma, los usuarios pueden agregar uno directamente en la vista del conjunto de datos apretando el botón de “Agregar Conjunto de Datos”.



Al seleccionar este botón se abre la vista que permite crear el conjunto de datos en cuestión:

🏠 / Conjuntos de datos / Crear Conjunto de datos

¿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.

1 Crear conjunto de datos

2 Agregar datos

Título:

ej. Un título descriptivo

* URL: datos.do.magnet.cl/dataset/<dataset> [Editar](#)

Descripción:

ej. Algunas notas útiles sobre los datos

Puede usar formato [Markdown](#) aquí

El primer paso consiste en ingresar la información del conjunto de datos en el sistema: título, descripción, fuente, visibilidad, organización, entre otros. Luego de que se tiene esto relleno hay que seleccionar “Siguiente: Agregar Datos” de esta forma se termina el primer paso y se continúa con el segundo. En el segundo paso de “Agregar datos” se muestra una pestaña para crear un recurso asociado al conjunto de datos ingresado anteriormente.

1 Crear conjunto de datos

2 Agregar datos

Fichero: [Subir](#) [Enlace](#)

Nombre: ej. Precios del Oro Enero 2011

Descripción: Algunas notas útiles sobre los datos

Puede usar formato [Markdown](#) aquí

Formato: ej. CSV, XML ó JSON

Esto será adivinado automáticamente. Déjelo en blanco si desea

[Anterior](#) [Guardar y Agregar otro](#) [Terminar](#)

Al seleccionar terminar se genera el conjunto de datos junto con los distintos recursos que se hayan creado en este punto. Una vez terminado se redirige al detalle del conjunto de datos ingresado.

Otra forma de agregar un recurso es a través del detalle de una organización, donde los editores de la organización pueden agregar un nuevo conjunto de datos y asociarlo a esta organización. Para ello deben seleccionar el botón “Administrar” en la organización e ir a la pestaña de “Conjunto de datos” en esta presionar el botón “Agrega conjunto de datos”:

Luego se sigue el mismo flujo anteriormente descrito.

Editar un conjunto de datos

Un conjunto de datos puede ser editado por los administradores y editores de la organización a la que pertenece el conjunto de datos y por el administrador del sistema. Para acceder a la edición de un conjunto de datos se debe seleccionar el botón “Administrar” que aparece en el detalle de este.

Este botón redirige a la vista de edición del conjunto de datos, en ella se pueden editar los metadatos del conjunto de datos y también los recursos asociados a este.

The screenshot shows the 'Portal de Datos Abiertos' interface. The top navigation bar includes 'Conjuntos de datos', 'Organizaciones', 'Grupos', and 'Acerca de', along with a search bar. The breadcrumb trail indicates the current location: 'Organizaciones / Armada de Chile / Reparticiones Navales más ... / Editar'. The main content area is titled 'Reparticiones Navales más importantes a lo largo del país' and shows 'Seguidores: 0'. The 'Editar metadatos' tab is active, displaying fields for 'Título' (filled with the dataset name), 'URL' (datos.do.magnet.cl/dataset/6927), and 'Descripción' (Dirección, ciudad y teléfono de contacto de las Reparticiones Navales más importantes a lo largo del país.). There is also a 'Ver conjunto de datos' button and a list of tags: Comandancia, Distrito, Estación, and Gobernación.

La edición de metadatos funciona de forma similar a la creación de un conjunto de datos. En la pestaña de Recursos se puede seleccionar el botón de “Añadir nuevo recurso” para agregar un nuevo recurso al conjunto de datos en cuestión. Además se encuentra la opción de reordenar los recursos del conjunto de datos para mostrarlos en el orden que se indiquen.

This screenshot shows the 'Recursos' tab of the same dataset. It features a '+ Añadir nuevo recurso' button and a 'Reorder resources' button. Below these, a single resource is listed with a green checkmark icon, the title 'Reparticiones Navales más importantes a lo ...', and the description 'Reparticiones Navales más importantes a lo largo del país.' The left sidebar remains the same as in the previous screenshot.

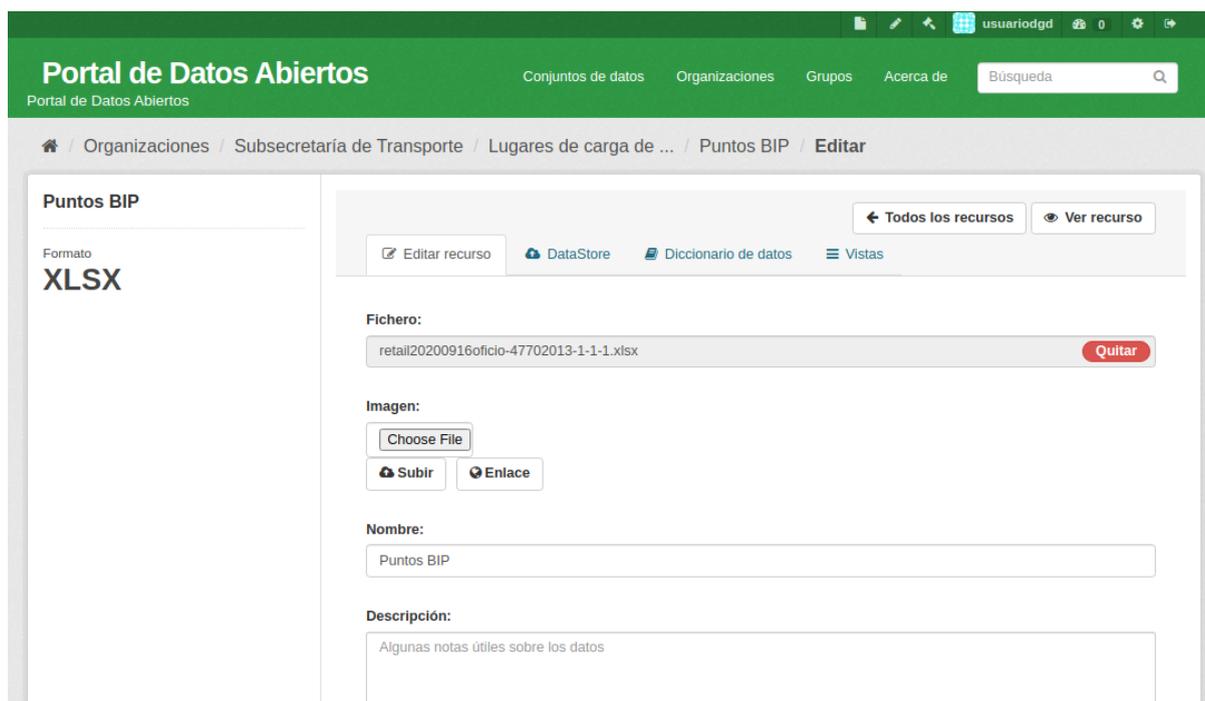
Si se aprieta en algún nombre de un recurso se accede a la vista de edición de este, descrita de mejor forma en la sección “Edición de un Recurso”.

Editar un Recurso

Para editar un recurso el usuario debe ser administrador o editor de la organización a la que pertenece el recurso o administrador del sistema. Para editar el recurso se debe seleccionar el botón "Administrar" en la vista de detalle del recurso.



Este botón redirige a la vista de administración del recurso la cual permite administrar la información del recurso, sus metadatos entre otros.



En "Editar recursos" se puede editar la información general del recursos, en "DataStore" se puede revisar el estado del recurso, si este fue cargado correctamente e intentar cargarlo nuevamente en caso de que sea necesario. En el diccionario de datos se puede editar la metadata de los campos del recurso, esta información es compartida al momento de descargar el recurso en algún formato en particular y al momento de consultar el recurso a través de la API.

Eliminar un Recurso

Para editar un recurso el usuario debe ser administrador o editor de la organización a la que pertenece el recurso o administrador del sistema. Para editar el recurso se debe seleccionar el botón “Administrar” en la vista de detalle del recurso.



datos.gob

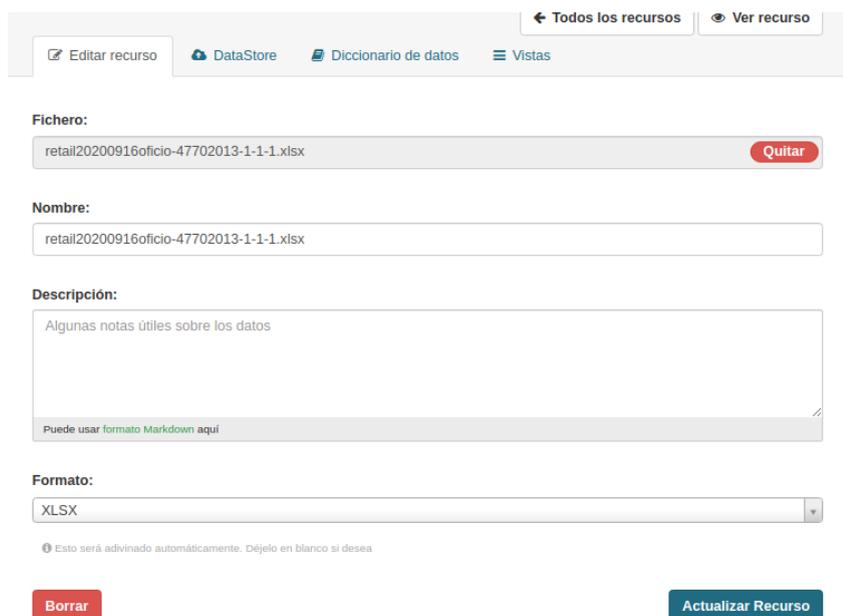
Conjuntos de Datos Organizaciones Categorías Documentos Ayuda Usuariodgd

Organizaciones > Agencia de Cooperación... > Conjunto de prueba > retail20200916oficio-477020 ...

retail20200916oficio-47702013-1-1-1.xlsx **Administrar**

<https://datos.do.magnet.cl/dataset/c9201c6a-61a2-4b57-957b-d84d37717a34/resource/d4d2aa0a-4a65-4f5a-a13d-07884542abff/download/retail20200916oficio-47702013-1-1-1.xlsx> **Descargar** API de datos

Luego en la vista de administración se debe seleccionar el botón “Borrar”.



← Todos los recursos Ver recurso

Editar recurso DataStore Diccionario de datos Vistas

Fichero: retail20200916oficio-47702013-1-1-1.xlsx **Quitar**

Nombre: retail20200916oficio-47702013-1-1-1.xlsx

Descripción: Algunas notas útiles sobre los datos
Puede usar formato Markdown aquí

Formato: XLSX
Esto será adivinado automáticamente. Déjelo en blanco si desea

Borrar **Actualizar Recurso**

Cuando se seleccione este botón aparecerá un pop-up advirtiendo de que se eliminará el recurso seleccionado, si se aprieta el botón “Confirmar” se eliminará el recurso del sistema.

Proceso de publicación utilizando API

La API de CKAN es una poderosa API, estilo RCP, que permite exponer las principales funcionalidades de CKAN. Desde esta API se pueden realizar múltiples acciones como por ejemplo:

- Obtener la lista de conjuntos de datos, categorías y otros objetos del sitio
- Obtener los conjuntos de datos y recursos en distintos formatos
- Buscar dentro de los conjuntos de datos
- Crear, actualizar y eliminar conjuntos de datos, recursos y otros objetos
- Obtener el flujo de actividad de los cambios que ha experimentado un recurso
- entre otros.

Se recomienda que al realizar la integración con la API se especifique la versión que se utilizará, en caso de que no se especifique la versión se utiliza la última, esto para evitar problemas con futuras actualizaciones del portal.

Para las llamadas que requieren autenticación se debe usar el token del usuario, este token se puede obtener en el perfil del usuario, en "Clave API" en la parte inferior del costado izquierdo. Los administradores pueden regenerar el API-token en caso de que se detecte un mal uso o el robo de este.



Acá se presentan algunas de las formas más comunes de consultar la API de datos del sitio, se puede ver mayor detalle en la [documentación del DataStore](#):

- Crear: https://datos.gob.cl/api/action/datastore_create
- Actualizar/Insertar: https://datos.gob.cl/api/action/datastore_upsert
- Consulta: https://datos.gob.cl/api/action/datastore_search
- Consulta (vía SQL): https://datos.gob.cl/api/action/datastore_search_sql

Autenticación para la API

La API utiliza el mismo sistema de autorización que la interfaz gráfica, es decir si existe alguna acción que el usuario puede realizar en la plataforma también la podrá realizar a través de la API. Esto quiere decir que todas las acciones que realiza un usuario autenticado en el sitio la tiene que realizar autenticándose con el API token a través de la API.

Para autenticarse se tiene que proveer la API key en la solicitud HTTP, para ello se debe incluir un header Authorization o X-CKAN-API-Key. Por ejemplo, si quisiéramos ver la lista de actividades del dashboard de usuario podríamos correr el siguiente código Python.

```
request = urllib2.Request('http://demo.ckan.org/api/3/action/dashboard_activity_list')
request.add_header('Authorization', 'API-TOKEN')
response_dict = json.loads(urllib2.urlopen(request, '{}').read())
```

Donde 'API-TOKEN' corresponde al token del usuario.

Funciones que permiten método GET

Las funciones definidas que permiten utilizar el método GET están definidas en [la documentación de la API](#), pueden ser llamadas utilizando el método GET. Por ejemplo, para ver la lista de datasets del sitio se puede llamar a la URL:

https://datos.gob.cl/api/3/action/package_list

Al ser una llamada get esta puede ser realizada desde un browser y se pueden ver los resultados.

Las llamadas GET permiten agregar distintos parámetros, para ello se debe agregar el símbolo '?' y luego los parámetros que se quieren definir separados por el símbolo '&'. Por ejemplo, si quisiéramos buscar 10 datasets que contengan la palabra gasto tendríamos que realizar la siguiente llamada:

https://datos.gob.cl/api/3/action/package_search?rows=10&q=gastos

Cuando el parámetro requiere una lista de parámetros strings, el valor se puede enviar agregando el parámetro múltiples veces en la URL.

Consultas de ejemplo:

Para obtener una lista con todas las etiquetas:

- En el browser: http://demo.ckan.org/api/3/action/tag_list
- Utilizando curl: `curl http://demo.ckan.org/api/3/action/tag_list`
- Utilizando ckanapi: `ckanapi -r http://demo.ckan.org action tag_list`

Administración de recursos utilizando la API

La API de CKAN permite administrar los distintos recursos, como se comentó en el punto anterior los permisos de los usuarios se mantienen entre el sitio web y la API por lo que para

realizar una acción en la que el usuario se necesita encontrar autenticado en el sitio se deberá también autenticar a través de la API.

Consulta de datos de un recurso

Para consultar datos de un recurso se debe realizar una consulta utilizando el método GET a url https://datos.gob.cl/api/action/datastore_search. Esta url permite recibir al menos los siguientes parámetros:

- **resource_id:** Identificador del recurso al cual se desea acceder. Este se puede obtener de la tabla "Información Adicional", campo "id". Este campo es requerido.
- **limit:** Permite seleccionar un límite de datos a retornar. (opcional)
- **offset:** Permite definir el offset de los datos que se consultan. (opcional)
- **q:** Permite realizar una búsqueda de texto en los distintos campos del recurso y retornar los que contengan este valor.
- **records_format:** Permite seleccionar el formato en que se quieren recibir los datos, por defecto es json, pero pueden ser 'objects', 'lists', 'csv' y 'tsv'.

Por ejemplo si quisiéramos los primeros 10 resultados del recurso con id 2d177f41-08f9-471a-af5c-bc949267f053 en formato csv debemos realizar la siguiente consulta:

- https://datos.gob.cl/api/action/datastore_search?resource_id=2d177f41-08f9-471a-af5c-bc949267f053&limit=10&records_format=csv

Si se quiere acceder a un conjunto de datos privados entonces el usuario se deberá encontrar autenticado.

Creación de un Conjunto de datos

Para crear un conjunto de datos existe la acción `package_create`. Para ello es necesario realizar POST a la url https://datos.gob.cl/api/action/package_create, a este endpoint se le debe enviar un json con la siguiente información:

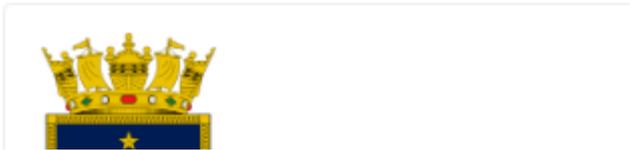
- **name:** nombre del conjunto de datos que se creara.
- **notes:** descripción larga del conjunto de datos que se creara
- **owner_org:** id o nombre de la organización en donde se cargará el conjunto de datos

Al ser una acción que requiere autenticación se debe agregar el header correspondiente en el request. El id o nombre de la organización se puede obtener desde la url de una organización como por ejemplo el id de la organización "Armada de Chile" será `armada_de_chile`:



[Organizaciones](#) ▶ Armada de Chile

Armada de Chile



Un request de ejemplo para python seria:

```
import http.client
import mimetypes
conn = http.client.HTTPSConnection("datos.gob.cl")
payload = "{\n  \"name\": \"data_set_de_prueba_api\",\n  \"notes\": \"A long description of my dataset\",\n  \"owner_org\": \"armada_de_chile\"}"
headers = {
  'Authorization': 'API-TOKEN',
  'Content-Type': 'text/plain'
}
conn.request("POST", "/api/action/package_create", payload, headers)
res = conn.getresponse()
data = res.read()
print(data.decode("utf-8"))
```

Donde 'API-TOKEN' corresponde al token del usuario.

Creación de un Recurso

Para crear un recurso existe la acción `datastore_create`, esta acción permite crear un recurso tanto en CKAN como en el DataStore. Para utilizar este método se debe hacer POST a la siguiente url: https://datos.gob.cl/api/3/action/datastore_create, se debe enviar un json con la siguiente información:

- `package_id`: id del conjunto de datos donde se quiere agregar el recurso.

- `resource_id`: se debe omitir en caso de que se desee crear un nuevo recurso en caso, si se agrega un id válido se reemplaza el recurso en cuestión.
- `resource`: diccionario con los datos del recurso que quiere ser cargado.
- `fields`: Lista o strings separados por coma de los distintos campos/columnas y su metadata adicional. (opcional)
- `records`: Datos que irán almacenados en los campos/columnas (opcional)
- `primary_key`: campos que representan una llave única (opcional).

Si se ingresa un `resource_id` entonces se actualizarán los campos y datos que se envíen en el request. Si no se ingresa un `resource_id` entonces se debe ingresar un `resource` con el recurso que será cargado por primera vez en el sistema.

A continuación se presenta un ejemplo en Python para crear un recurso en el conjunto de datos anteriormente creado (`data_set_de_prueba_api`). El campo `resource` debe ser un diccionario que debe seguir la siguiente estructura:

- `package_id`: conjunto de datos donde se debe agregar el recurso.
- `url`: url del recurso que se cargará
- `revision_id` (string) (opcional)
- `description`: descripción del recurso (opcional)
- `format` (string): formato del recurso (opcional)
- `hash` (string) (opcional)
- `name` (string): nombre del recurso (opcional)
- `resource_type` (string) (opcional)
- `cache_url` (string) (opcional)
- `mime type` (string) (opcional)
- `mimetype_inner` (string) (opcional)
- `cache_url` (string) (opcional)
- `size` (int) (opcional)
- `created` (iso date string) (opcional)
- `last_modified` (iso date string) (opcional)
- `cache_last_updated` (iso date string) (opcional)
- `upload` (FieldStorage (optional) needs multipart/form-data) (opcional)

Llamada de ejemplo¹ en Python:

```
import http.client
import mimetypes
conn = http.client.HTTPSConnection("datos.gob.cl")
payload = "{\n  \"package_id\": \"data_set_de_prueba_api\",\n  \"resource\": {\n    \"package_id\": \"data_set_de_prueba_api\",\n    \"url\":\n    \"https://raw.githubusercontent.com/MinCiencia/Datos-COVID19/master/output/producto1
```

¹ Para el siguiente ejemplo se utilizan los datos puestos a disposición por el Ministerio de Ciencia, Conocimiento, Tecnología e Innovación en su cuenta <https://github.com/MinCiencia/Datos-COVID19>

```
9/CasosActivosPorComuna.csv",\n    \ "description": \"recurso para ejemplificar carga de
datos utilizando la API\", \n    \ "name": \"Recurso de prueba\" \n } \n }"
headers = {
'Authorization': 'API-TOKEN',
'Content-Type': 'text/plain'
}
conn.request("POST", "/api/3/action/datastore_create", payload, headers)
res = conn.getresponse()
data = res.read()
print(data.decode("utf-8"))
```

Para mayor información de esta acción revisar la documentación en:

https://docs.ckan.org/en/ckan-2.7.3/maintaining/datastore.html#ckanext.datastore.logic.action.datastore_create y

https://docs.ckan.org/en/2.8/api/index.html#ckan.logic.action.create.resource_create

Edición/inserción de datos en un Recurso

Para editar o insertar datos en un recurso existe la acción `datastore_upsert_`. Para ello es necesario realizar POST a la url https://datos.gob.cl/api/action/datastore_upsert, a este endpoint se le debe enviar un json con la siguiente información:

- `resource_id` (string): identificador del recurso.
- `force` (booleano): Setear como True si se desea editar un recurso que es solo lectura.
- `records` (lista de diccionarios): los datos, por ejemplo [{"nombre": "Juan", "apellido": "Perez"}, {"nombre": "Jorge", "apellido": "Gonzales"}]
- `method` (string): el método con el cual se usará el put en el DataStore. Las opciones posibles son: `upsert`, `insert`, `update`. Por defecto se selecciona `upsert`.

Características de cada uno de los métodos:

- `upsert`: Actualiza el dato si existe uno con la misma llave primaria, en caso de que no exista lo inserta. Requiere que el recurso tenga seteada una llave primaria. (se puede setear con el método `datastore_create`).
- `insert`: Este método es más rápido que el `upsert`, pero fallará si el dato insertado concuerda con uno insertado previamente. No requiere que se encuentre seteada una llave primaria.
- `update`: Solo actualiza los datos existentes. Ocurrirá una excepción si no existe un dato que se quiera actualizar. Requiere que exista una llave primaria (única).

Por ejemplo si quisiéramos insertar y actualizar alguno de los datos ingresados en el recurso. Para ello lo primero que tenemos que hacer es setear una llave primaria, a continuación una llamada en Python que hace esto:

```
import http.client
import mimetypes
```

```

conn = http.client.HTTPSConnection("datos.gob.cl")
payload = "{\n  \"package_id\": \"data_set_de_prueba_api\",\n  \"resource_id\":\n  \"97aadb0-468d-4289-85ae-b0f0b4f04c95\",\n  \"primary_key\": [\"Codigo comuna\"],\n  \"force\": \"True\"\n}"
headers = {
  'Authorization': 'API-TOKEN',
  'Content-Type': 'text/plain'
}
conn.request("POST", "/api/3/action/datastore_create", payload, headers)
res = conn.getresponse()
data = res.read()
print(data.decode("utf-8"))

```

Donde 'API-TOKEN' corresponde al token del usuario.

Luego podemos hacer la llamada con el método upsert para actualizar los datos del recurso, a continuación la llamada de ejemplo en Python:

```

import http.client
import mimetypes
conn = http.client.HTTPSConnection("datos.gob.cl")
payload = "{\n  \"resource_id\": \"97aadb0-468d-4289-85ae-b0f0b4f04c95\",\n  \"force\":\n  \"True\",\n  \"records\": [{\n    \"Region\": \"Region de prueba\",\n    \"Codigo region\": \"20\",\n    \"Comuna\": \"Arica\",\n    \"Codigo comuna\":\n    \"151012\",\n    \"Poblacion\": \"247552.0\",\n    \"2020-04-13\": \"88.0\",\n    \"2020-04-15\": \"80.0\",\n    \"2020-04-17\": \"72.0\",\n    \"2020-04-20\":\n    \"85.0\",\n    \"2020-04-24\": \"77.0\",\n    \"2020-04-27\": \"103.0\",\n    \"2020-05-01\": \"105.0\",\n    \"2020-05-04\": \"87.0\",\n    \"2020-05-08\":\n    \"61.0\",\n    \"2020-05-11\": \"66.0\",\n    \"2020-05-15\": \"49.0\",\n    \"2020-05-18\": \"68.0\",\n    \"2020-05-22\": \"117.0\",\n    \"2020-05-25\":\n    \"143.0\",\n    \"2020-05-29\": \"163.0\",\n    \"2020-06-01\": \"177.0\",\n    \"2020-06-05\": \"248.0\",\n    \"2020-06-08\": \"278.0\",\n    \"2020-06-12\":\n    \"328.0\",\n    \"2020-06-15\": \"353.0\",\n    \"2020-06-19\": \"358.0\",\n    \"2020-06-23\": \"381.0\",\n    \"2020-06-28\": \"408.0\",\n    \"2020-07-01\":\n    \"406.0\",\n    \"2020-07-05\": \"495.0\",\n    \"2020-07-10\": \"575.0\",\n    \"2020-07-13\": \"655.0\",\n    \"2020-07-17\": \"818.0\",\n    \"2020-07-20\":\n    \"859.0\",\n    \"2020-07-24\": \"981.0\",\n    \"2020-07-27\": \"984.0\",\n    \"2020-07-31\": \"962.0\",\n    \"2020-08-03\": \"861.0\",\n    \"2020-08-07\":\n    \"709.0\",\n    \"2020-08-10\": \"779.0\",\n    \"2020-08-14\": \"775.0\",\n    \"2020-08-17\": \"778.0\",\n    \"2020-08-21\": \"651.0\",\n    \"2020-08-24\":\n    \"643.0\",\n    \"2020-08-28\": \"585.0\",\n    \"2020-08-31\": \"592.0\",\n    \"2020-09-04\": \"529.0\",\n    \"2020-09-07\": \"565.0\",\n    \"2020-09-11\":\n    \"486.0\",\n    \"2020-09-14\": \"468.0\",\n    \"2020-09-18\": \"390.0\",\n    \"2020-09-21\": \"402.0\",\n    \"2020-09-25\": \"363.0\",\n    \"2020-09-28\":\n    \"399.0\",\n    \"2020-10-02\": \"432.0\",\n    \"2020-10-05\": \"515.0\",\n    \"2020-10-09\": \"561.0\",\n    \"2020-10-12\": \"552.0\",\n    \"2020-10-16\":\n    \"506.0\",\n    \"2020-10-19\": \"485.0\",\n    \"2020-10-24\": \"319.0\",\n    \"2020-10-26\": \"316.0\",\n    \"2020-10-30\": \"269.0\",\n    \"2020-11-02\":\n    \"226.0\"\n  }],\n  \"Region\": \"Arica y Parinacota\",\n  \"Codigo region\": \"15\",\n  \"Comuna\": \"Camarones\",\n  \"Codigo

```

```

comuna\": \"15102\", \n      \"Poblacion\": \"1233.0\", \n      \"2020-04-13\":
\"0.0\", \n      \"2020-04-15\": \"0.0\", \n      \"2020-04-17\": \"0.0\", \n
\"2020-04-20\": \"0.0\", \n      \"2020-04-24\": \"0.0\", \n      \"2020-04-27\": \"0.0\", \n
\"2020-05-01\": \"0.0\", \n      \"2020-05-04\": \"0.0\", \n      \"2020-05-08\": \"0.0\", \n
\"2020-05-11\": \"0.0\", \n      \"2020-05-15\": \"0.0\", \n      \"2020-05-18\": \"0.0\", \n
\"2020-05-22\": \"10.0\", \n      \"2020-05-25\": \"10.0\", \n      \"2020-05-29\":
\"0.0\", \n      \"2020-06-01\": \"0.0\", \n      \"2020-06-05\": \"0.0\", \n
\"2020-06-08\": \"0.0\", \n      \"2020-06-12\": \"0.0\", \n      \"2020-06-15\": \"0.0\", \n
\"2020-06-19\": \"0.0\", \n      \"2020-06-23\": \"0.0\", \n      \"2020-06-28\": \"0.0\", \n
\"2020-07-01\": \"0.0\", \n      \"2020-07-05\": \"0.0\", \n      \"2020-07-10\": \"0.0\", \n
\"2020-07-13\": \"8.0\", \n      \"2020-07-17\": \"12.0\", \n      \"2020-07-20\":
\"12.0\", \n      \"2020-07-24\": \"7.0\", \n      \"2020-07-27\": \"2.0\", \n
\"2020-07-31\": \"2.0\", \n      \"2020-08-03\": \"6.0\", \n      \"2020-08-07\": \"4.0\", \n
\"2020-08-10\": \"4.0\", \n      \"2020-08-14\": \"1.0\", \n      \"2020-08-17\": \"1.0\", \n
\"2020-08-21\": \"1.0\", \n      \"2020-08-24\": \"3.0\", \n      \"2020-08-28\": \"0.0\", \n
\"2020-08-31\": \"1.0\", \n      \"2020-09-04\": \"1.0\", \n      \"2020-09-07\": \"1.0\", \n
\"2020-09-11\": \"0.0\", \n      \"2020-09-14\": \"0.0\", \n      \"2020-09-18\": \"0.0\", \n
\"2020-09-21\": \"1.0\", \n      \"2020-09-25\": \"0.0\", \n      \"2020-09-28\": \"0.0\", \n
\"2020-10-02\": \"0.0\", \n      \"2020-10-05\": \"1.0\", \n      \"2020-10-09\": \"1.0\", \n
\"2020-10-12\": \"1.0\", \n      \"2020-10-16\": \"1.0\", \n      \"2020-10-19\": \"0.0\", \n
\"2020-10-24\": \"0.0\", \n      \"2020-10-26\": \"0.0\", \n      \"2020-10-30\": \"0.0\", \n
\"2020-11-02\": \"0.0\" \n    }], \n    \"method\": \"upsert\" \n  }
headers = {
  'Authorization': 'API-TOKEN',
  'Content-Type': 'text/plain'
}
conn.request(\"POST\", \"/api/3/action/datastore_upsert\", payload, headers)
res = conn.getresponse()
data = res.read()
print(data.decode(\"utf-8\"))

```

Donde 'API-TOKEN' corresponde al token del usuario.

Para un mayor detalle sobre los distintos métodos se recomienda revisar la documentación oficial.

Proceso de publicación usando Pentaho en el portal de datos

Pentaho es una herramienta externa a CKAN que ayuda con el análisis y manipulación de datos.

Existe el Pentaho Data Integration (PDI) que se instala de forma local en la computadora de un usuario. Esta herramienta le permite al usuario crear flujos para procesar y manipular datos de forma visual a través de steps configurables.

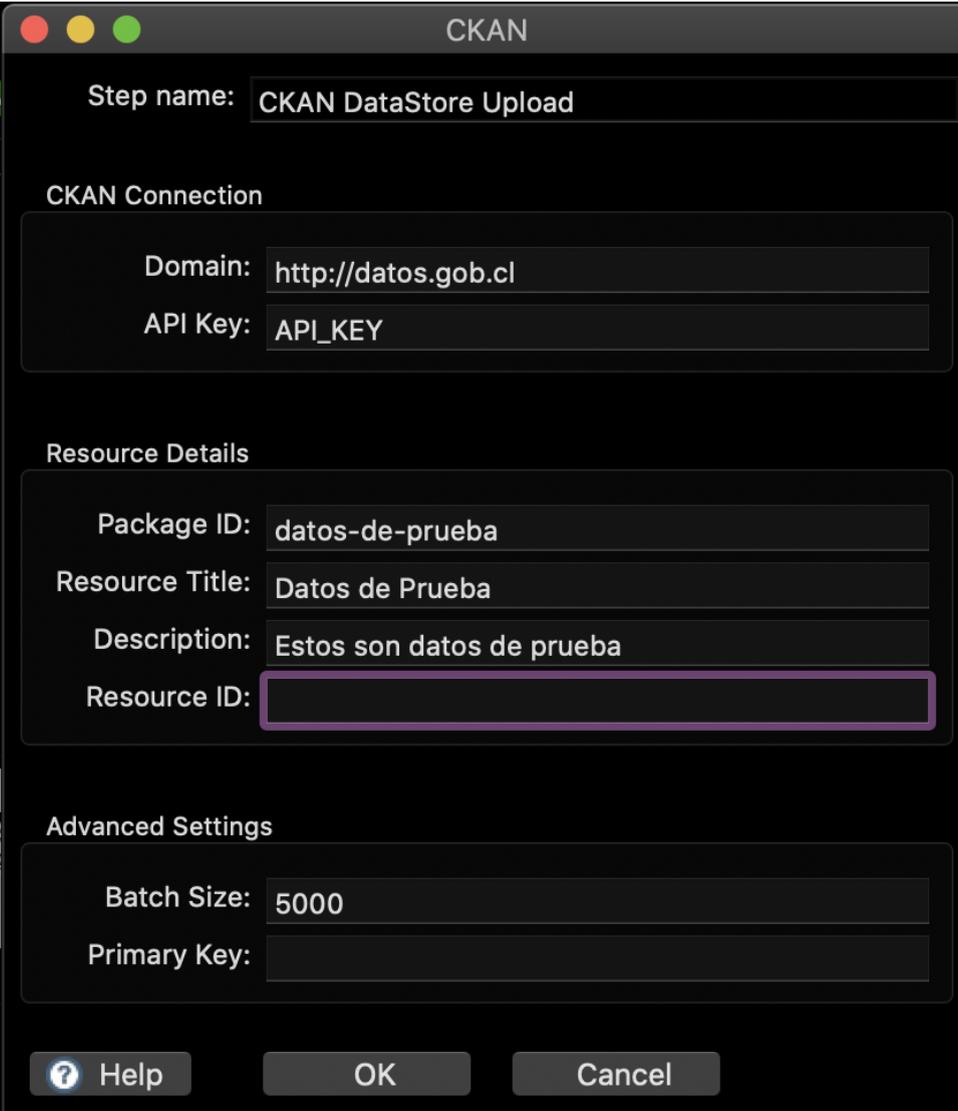
También existe el Pentaho Community Dashboard Editor (CDE) que permite crear gráficas y mantener un dashboard con información actualizada a partir de distintas fuentes de datos.

Esta herramienta permite conectarse con un API o utilizar directamente los datos resultantes de una transformación hecha en el PDI, para luego utilizarla en la construcción de gráficos.

Subir datos usando Pentaho (PDI)

Es posible usar el PDI para subir recursos directamente a CKAN. Para esto es necesario instalar la extensión CKAN DataStore Upload que se encuentra en el marketplace de extensiones del PDI. Para acceder al marketplace desde el menú seleccionar tools, luego marketplace, buscar CKAN DataStore Upload y finalmente instalar la extensión. Alternativamente a esta extensión también es posible utilizar los steps REST Client o HTTP Client para interactuar directamente con la API.

Una vez esté instalada la extensión el step CKAN DataStore Upload estará disponible. Esta extensión permitirá subir un recurso nuevo o actualizar uno existente dentro de un conjunto de datos, por lo que será necesario contar con la ID del conjunto de datos.



The image shows a configuration dialog for the 'CKAN DataStore Upload' step. The dialog is titled 'CKAN' and contains the following sections:

- Step name:** CKAN DataStore Upload
- CKAN Connection:**
 - Domain:
 - API Key:
- Resource Details:**
 - Package ID:
 - Resource Title:
 - Description:
 - Resource ID:
- Advanced Settings:**
 - Batch Size:
 - Primary Key:

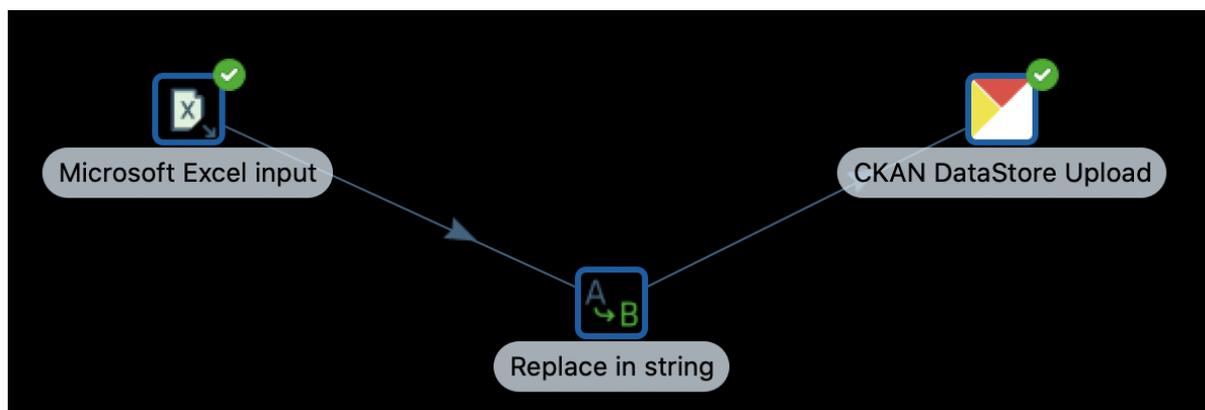
At the bottom, there are three buttons: 'Help' (with a question mark icon), 'OK', and 'Cancel'.

Los campos a configurar son:

- **Domain:** la url base de donde se obtendrán los datos (<https://datos.gob.cl>)
- **API Key:** la llave del usuario que subirá los datos, esta se obtiene desde el perfil del usuario (ver sección Consumo API).
- **Package ID:** la id del conjunto de datos al cual se agrega/actualiza el recurso, esta corresponde a la parte de la url del conjunto de datos que lo identifica ej: <https://datos.gob.cl/dataset/datos-de-prueba>
- **Resource Title:** el nombre que tendrá el recurso
- **Description:** la descripción que tendrá el recurso
- **Resource ID:** la id del recurso que se actualizará o dejar en blanco en caso de ser un recurso nuevo, esta corresponde a la parte de la url del recurso que lo identifica ej: <https://datos.gob.cl/dataset/datos-de-prueba/resource/recurso-de-prueba>
- **Batch Size:** la cantidad de filas que se enviarán en cada lote, por defecto es 5000
- **Primary Key:** en caso de especificar el nombre de las columnas (ej: campo1;campo2;campo3), la extensión actualizará esas columnas en lugar de solo insertar datos

Este step requiere que se le entreguen datos como input desde otro step, estos datos serán los que finalmente se suban a CKAN.

Ejemplo: Tengo un documento con datos tabulados que contienen la información de la comuna en la que residen un grupo de la personas, pero el dato de la comuna que proporcionó cada persona no se encuentra estandarizado, cada persona ingreso el nombre de su comuna de manera distinta (ej: Ñuñoaa, nunoa, ÑUnOa, etc). Es posible mediante el PDI realizar una transformación de datos que lea el documento y estandarice el nombre de todas las comunas y luego el resultado de esta transformación sea lo que se publique en datos.gob.cl para así facilitar el consumo de los datos por otras instituciones.



Primero tomamos un step Microsoft Excel Input, en el cual se agrega un archivo de excel. Luego el step Replace in String permite encontrar un texto y cambiarlo por otro. Finalmente el step CKAN DataStore Upload sube el resultado final de la transformación a datos.gob.cl.

