El futuro digital es de todos

MinTIC

OPEN REFINE

El futuro digital es de todos

MinTIC

REGLAS DEL TALLER

Horario
10:00 a.m.
11:00 a.m.

No abandonar la reunión

Presencia plena

Preguntas por medio del link o al final de la presentación

INTRODUCCIÓN

La falta de calidad de los datos es uno de los principales inconvenientes a los que se enfrentan los responsables de sistemas de información y las empresas en general, pues representa claramente uno de los problemas "ocultos" más graves y persistentes en cualquier organización.

En efecto, una buena calidad de datos es el activo corporativo más potente, ya que permite acelerar el crecimiento y administrar de mejor manera los costos y las iniciativas para obtener mejores rentabilidades.

El futuro digital es de todos

MinTIC

¿Qué es la iniciativa de software libre?

El futuro digital es de todos

MinTIC

Objetivos de la iniciativa de Software libre

El futuro digital es de todos

MinTIC

¿Qué es OPEN REFINE?

Es una herramienta que nos permite perfilar y limpiar conjuntos de datos.

CARACTERÍSTICAS

Es una aplicación de escritorio que se puede descargar e instalar desde la web : **http://openrefine.org/**
Está disponible para **Windows, Mac y Linux.**

Permite transformar archivos de diferentes extensiones como XLS, CSV, JSON, XML, TSV y las **spreadsheets** de Google.

Utiliza lenguaje GREL en el que podemos usar expresiones regulares, concretamente en sus funciones: **replace, match, partition, repartition y split.**

Es compatible con cualquier navegador.

El futuro digital es de todos | MinTIC

CONSEJOS DE ELECCIÓN DE UN LENGUAJE DE PROGRAMACIÓN Y RAZONES PARA APRENDER OPEN REFINE

Ámbito de aplicación.

Costo. El factor económico es también crucial en nuestra decisión

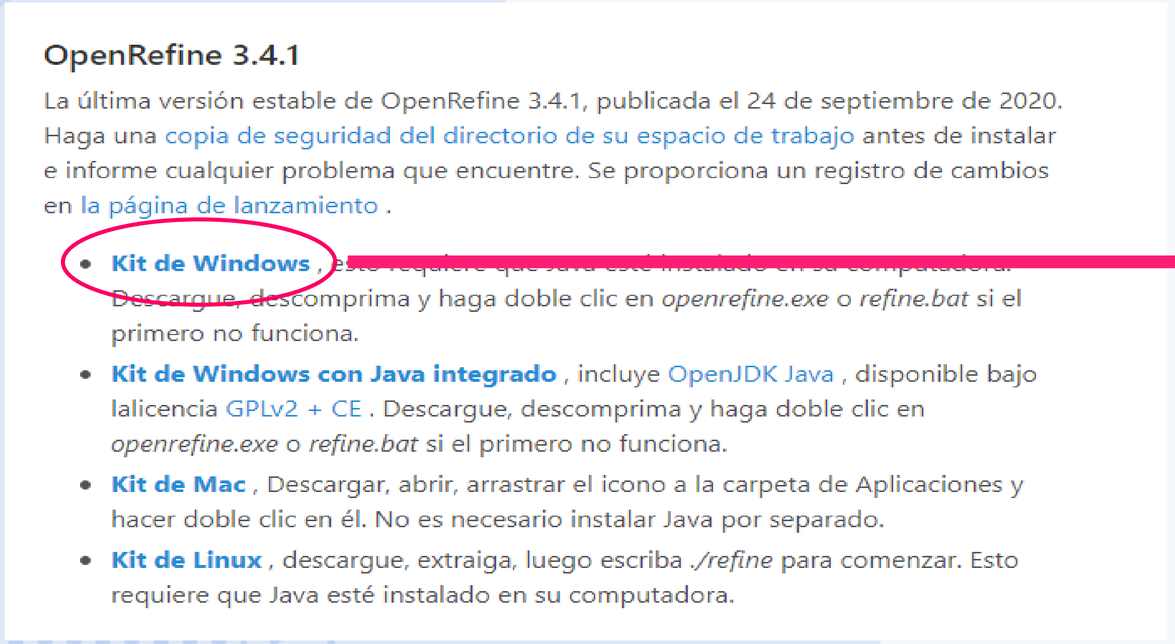
Características propias del lenguaje o entorno de programación.

El futuro digital es de todos | MinTIC

PROCEDIMIENTO RÁPIDO PARA LA DESCARGA E INSTALACIÓN DE LA HERRAMIENTA

Entramos a la página oficial de OPEN REFINE
y seleccionamos la opción DOWNLOAD - DESCARGAR

https://openrefine.org/download.html



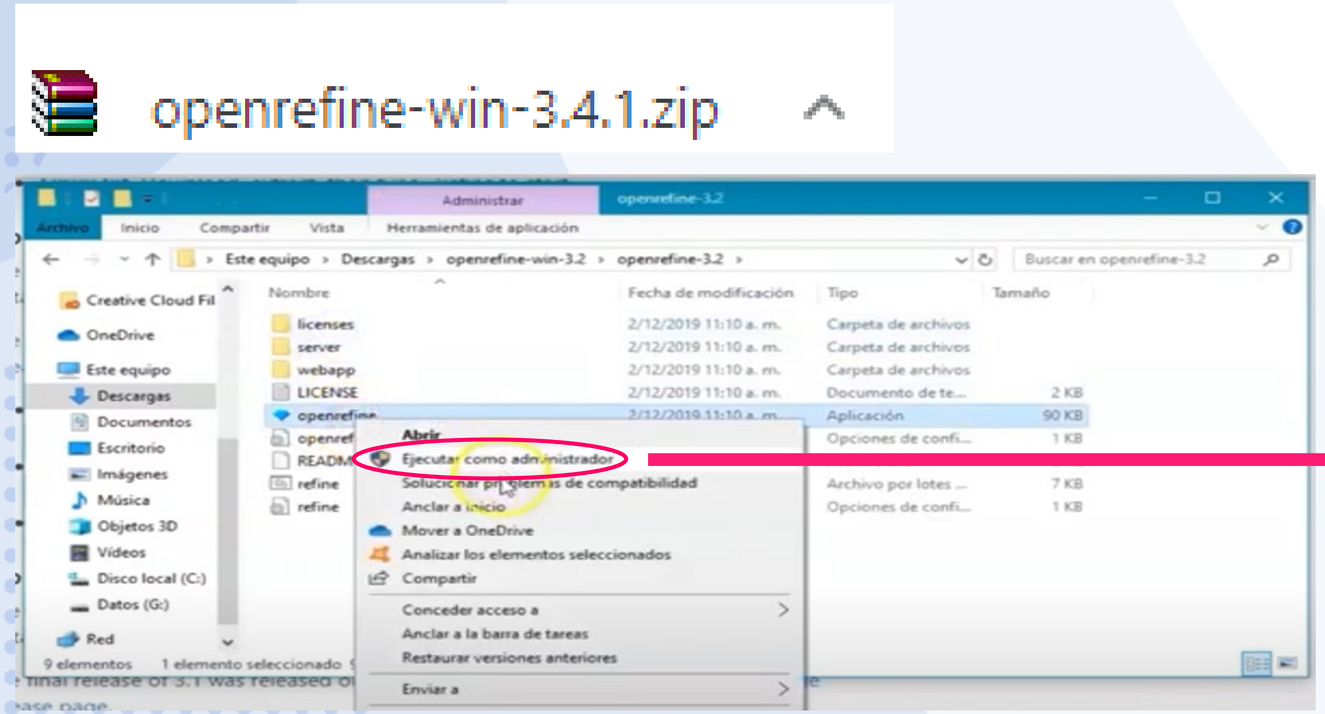
OpenRefine 3.4.1

La última versión estable de OpenRefine 3.4.1, publicada el 24 de septiembre de 2020. Haga una copia de seguridad del directorio de su espacio de trabajo antes de instalar e informe cualquier problema que encuentre. Se proporciona un registro de cambios en la página de lanzamiento .

- **Kit de Windows** , esto requiere que Java esté instalado en su computadora. Descargue, descomprima y haga doble clic en *openrefine.exe* o *refine.bat* si el primero no funciona.
- **Kit de Windows con Java integrado** , incluye OpenJDK Java , disponible bajo lalicencia GPLv2 + CE . Descargue, descomprima y haga doble clic en *openrefine.exe* o *refine.bat* si el primero no funciona.
- **Kit de Mac** , Descargar, abrir, arrastrar el icono a la carpeta de Aplicaciones y hacer doble clic en él. No es necesario instalar Java por separado.
- **Kit de Linux** , descargue, extraiga, luego escriba *./refine* para comenzar. Esto requiere que Java esté instalado en su computadora.

SELECCIONAR

Nos descarga una carpeta .ZIP para descomprimir

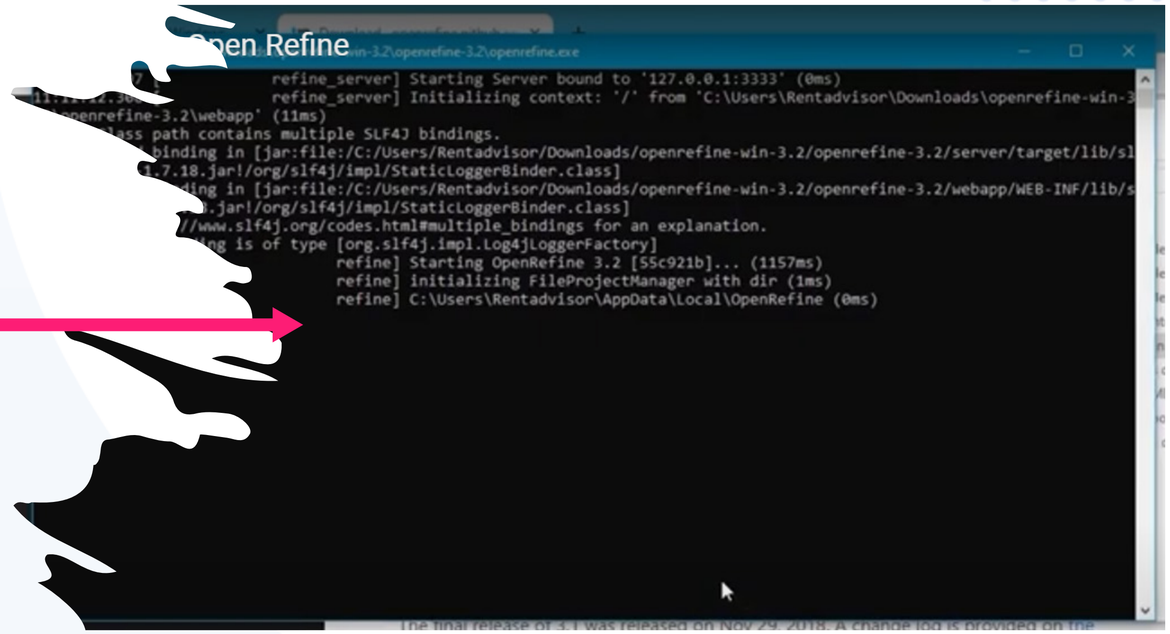


Vamos a la carpeta de descarga, seleccionamos la herramienta **OPEN REFINE,** y damos clic derecho a la opción **EJECUTAR COMO ADMINISTRADOR**

El futuro digital es de todos | MinTIC

MIENTRAS DESCARGA...

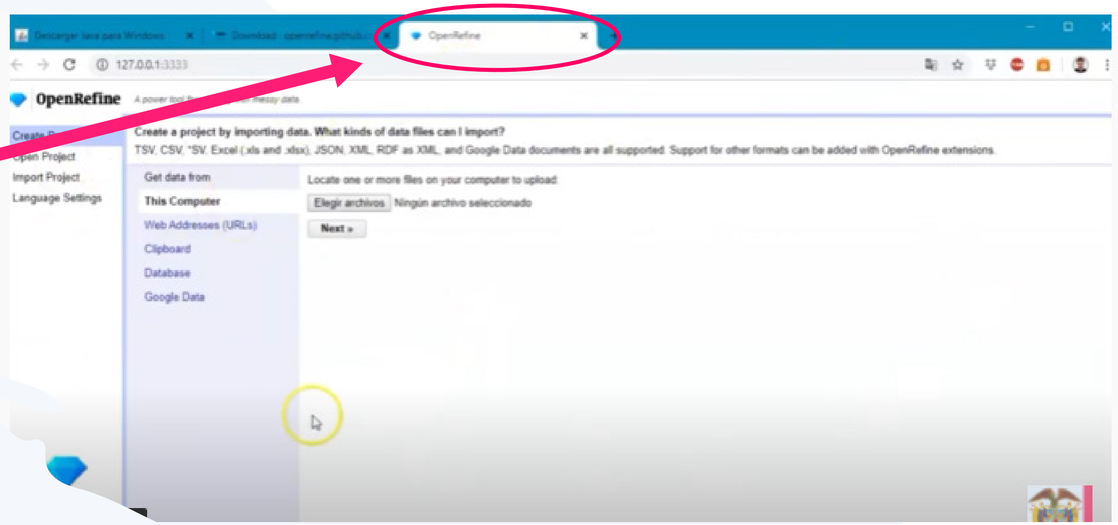
Nos muestra esta interfaz donde ya se está ejecutando el sistema



Automáticamente se abrirá una pestaña nueva en el en el navegador

CON LA SIGUIENTE DIRECCIÓN IP

http://127.0.0.1:3333/



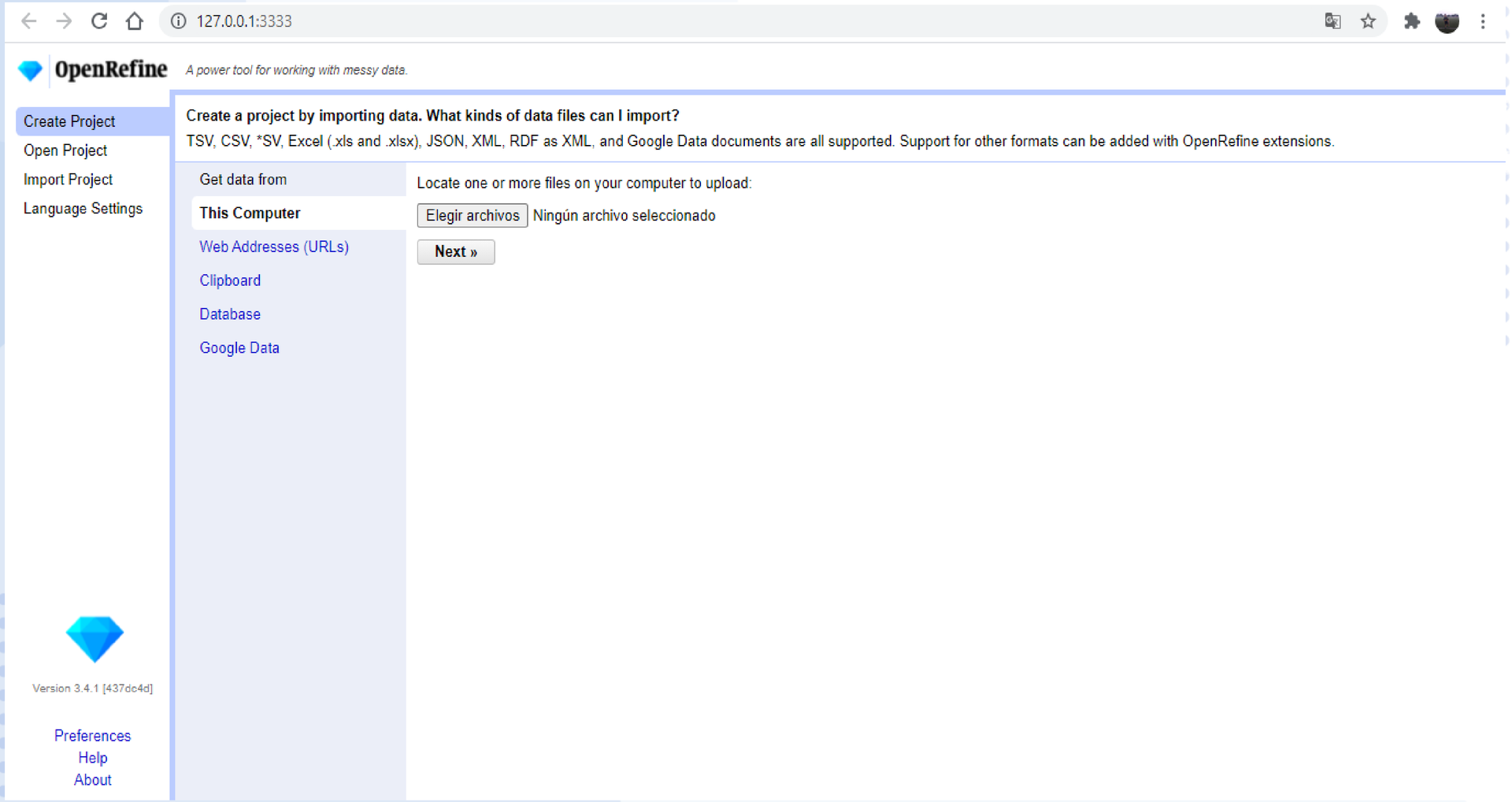
El futuro digital es de todos

MinTIC

EJEMPLO DE USO

El futuro digital es de todos

MinTIC



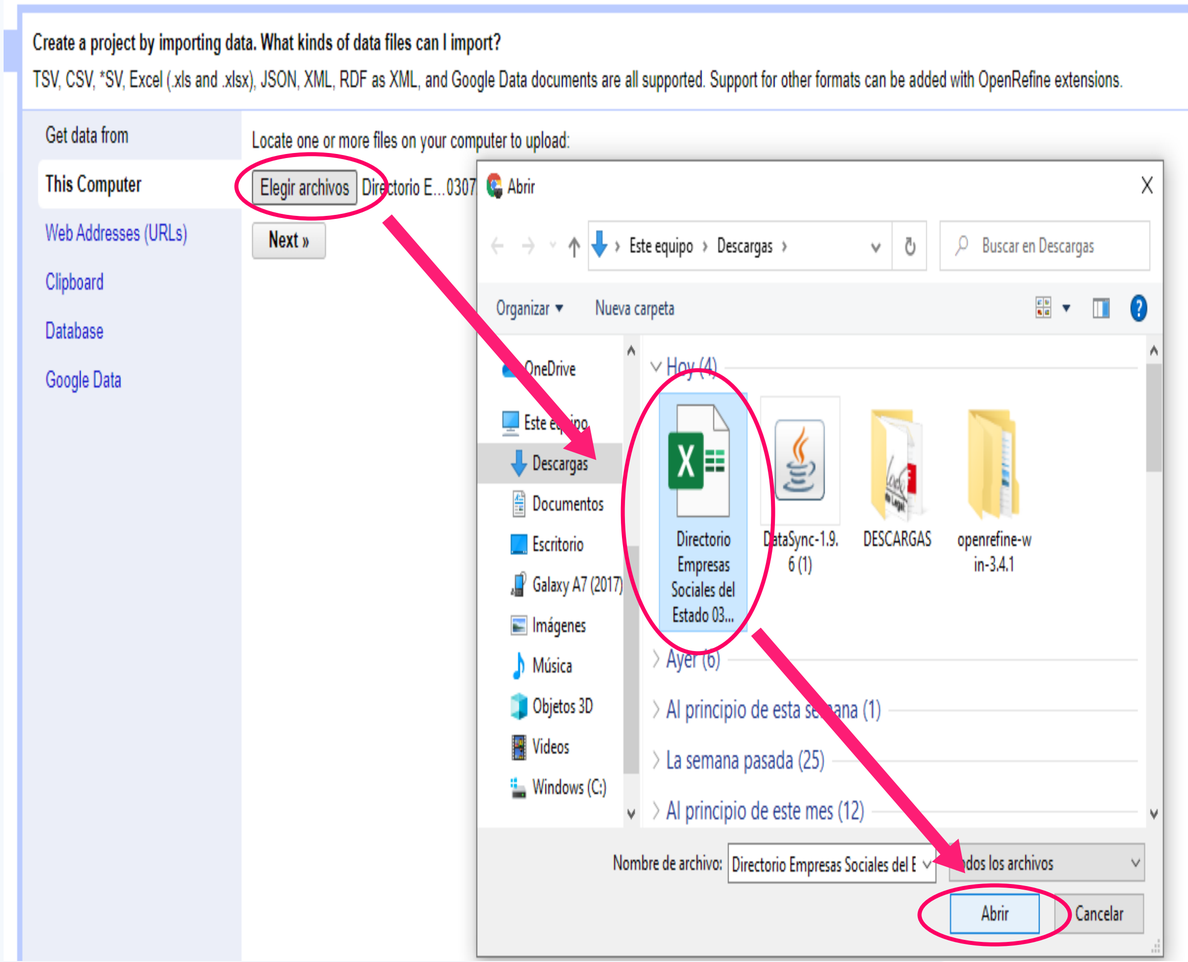
ENTRAMOS EN LA INTERFAZ DE OPEN REFINE

El futuro digital es de todos | MinTIC

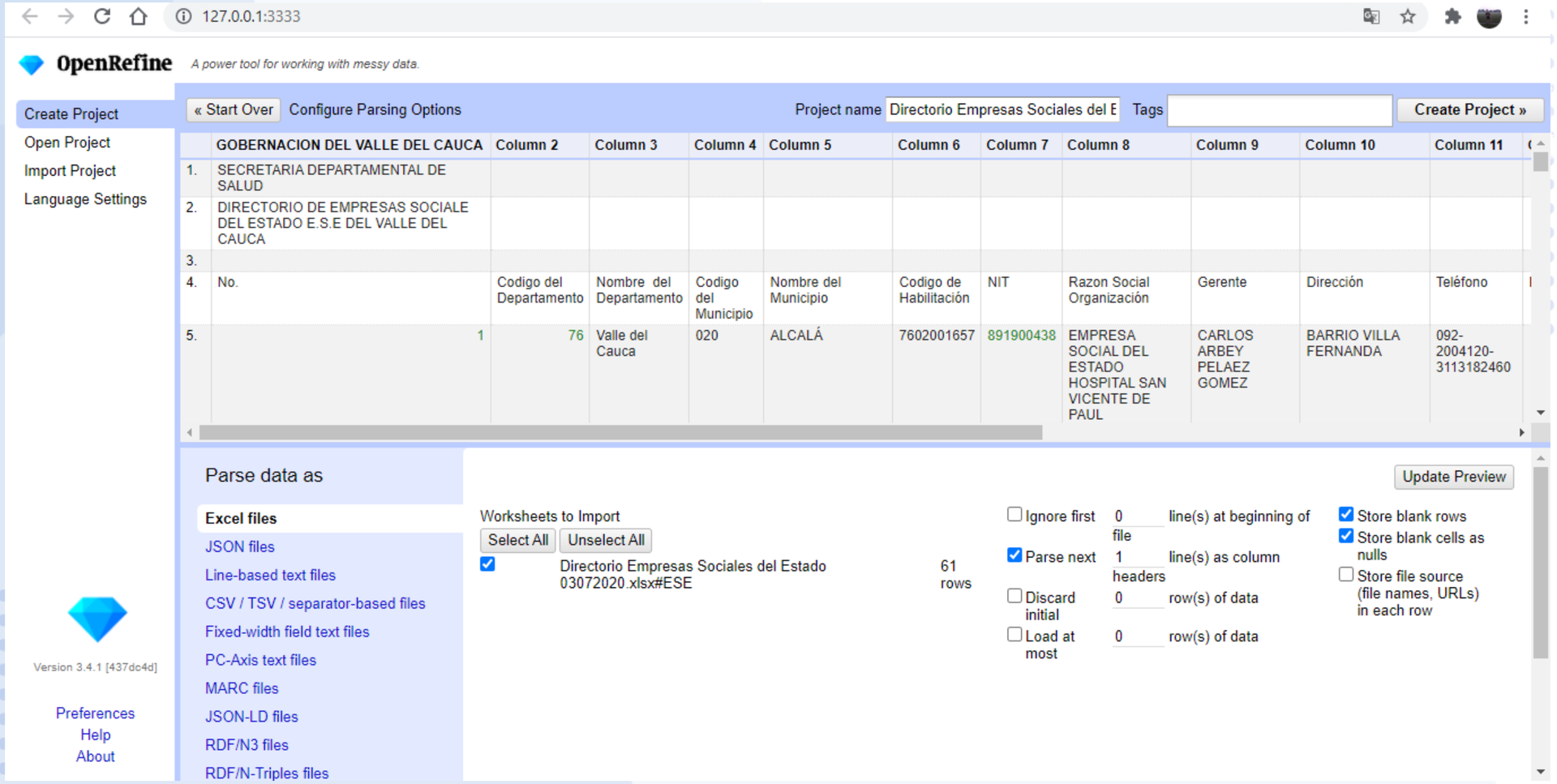
Seleccionamos

Elegir archivos

lo que nos lleva a la carpeta donde elegiremos el documento y le damos clic en **ABRIR**

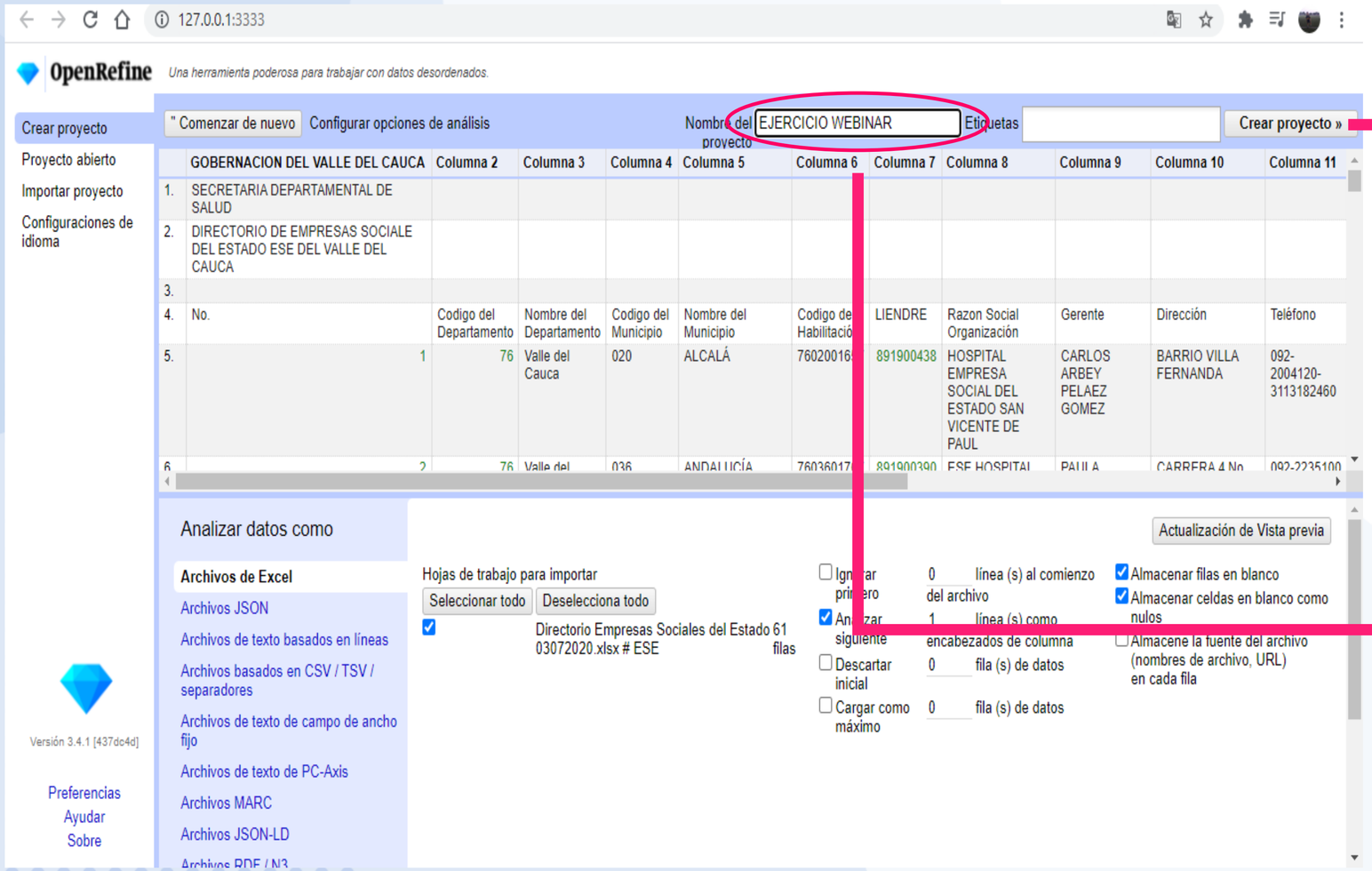


El futuro digital es de todos | MinTIC



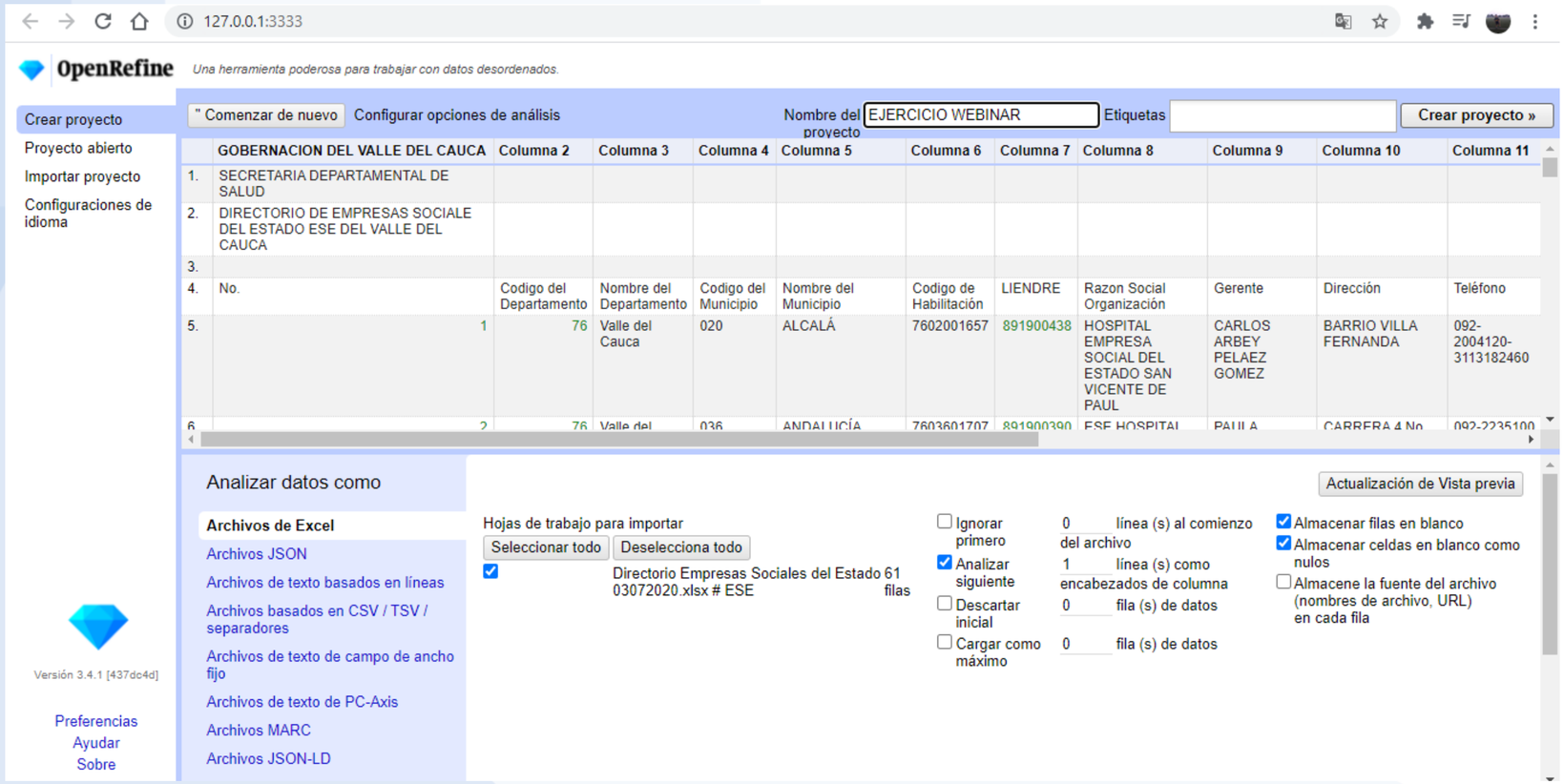
YA TENEMOS EL ARCHIVO EN NUESTRA INTERFAZ PARA TRABAJAR SOBRE ELLA

El futuro digital es de todos | MinTIC



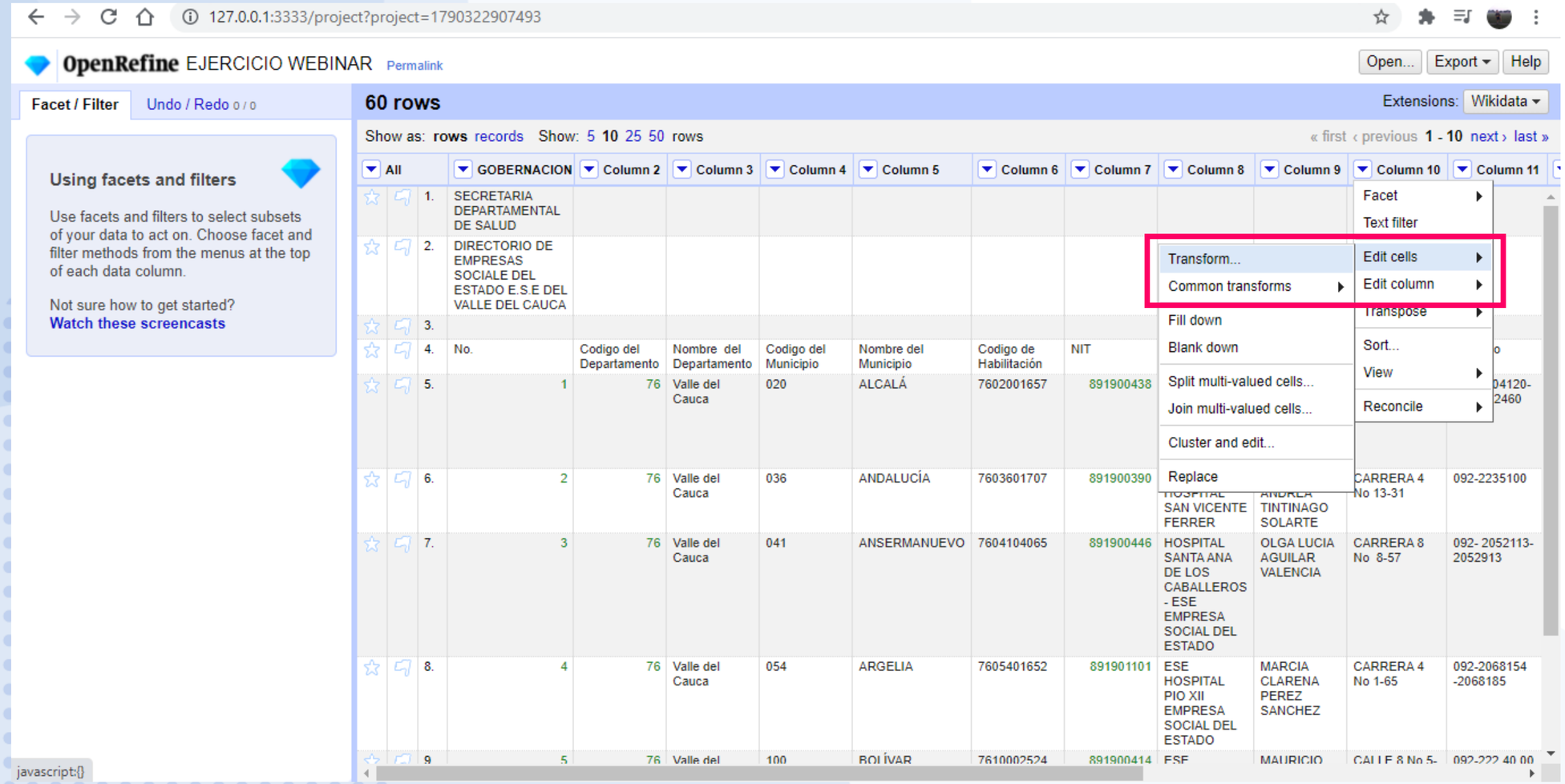
CREAR PROYECTO

Personalizamos el nombre del proyecto



Trabajando con funciones de GREL en Open Refine nos permite crear nuestras propias facetas ad hoc usando GREL, Jython o Clojure. GREL (Google Refine Expression Language) es un lenguaje de programación con un gran número de funciones que permite realizar tareas de depuración avanzadas. Veamos algún caso sencillo. Vamos a modificar las abreviaturas de los nombre de los viales a su denominación extendida. En el submenú de operaciones de la columna 'vial' seleccionaremos **Edits cells -› Transform** e ingresamos la siguiente expresión GREL: value.replace('Avdav', 'Avenida')

El futuro digital es de todos | MinTIC



Seleccionamos la opción

- **EDIT CELL**
- **TRANSFORM**

El futuro digital es de todos | MinTIC

Transformación de texto personalizado en la columna Columna 10

Expresión

```
value.replace('CRA','CARRERA')
```

Idioma: Lenguaje de expresión general refinado (GREL)

Sin error de sintaxis.

Avance | Historia | Sembrado de estrellas | Ayudar

3.	nulo	Error: el reemplazo espera 3 cadenas o 1 cadena, 1 expresión regular y 1 cadena
4.	Dirección	Dirección
5.	BARRIO VILLA FERNANDA	BARRIO VILLA FERNANDA
6.	CRA 4 No 13-31	CARRERA 4 No 13-31
7.	CARRERA 8 No 8-57	CARRERA 8 No 8-57
8.	CRA 4 No 1-65	CARRERA 4 No 1-65
9.	CALLE 8 No 5-09	CALLE 8 No 5-09
10	CRA 47 No 2-08	CARRERA 47 No 2-08

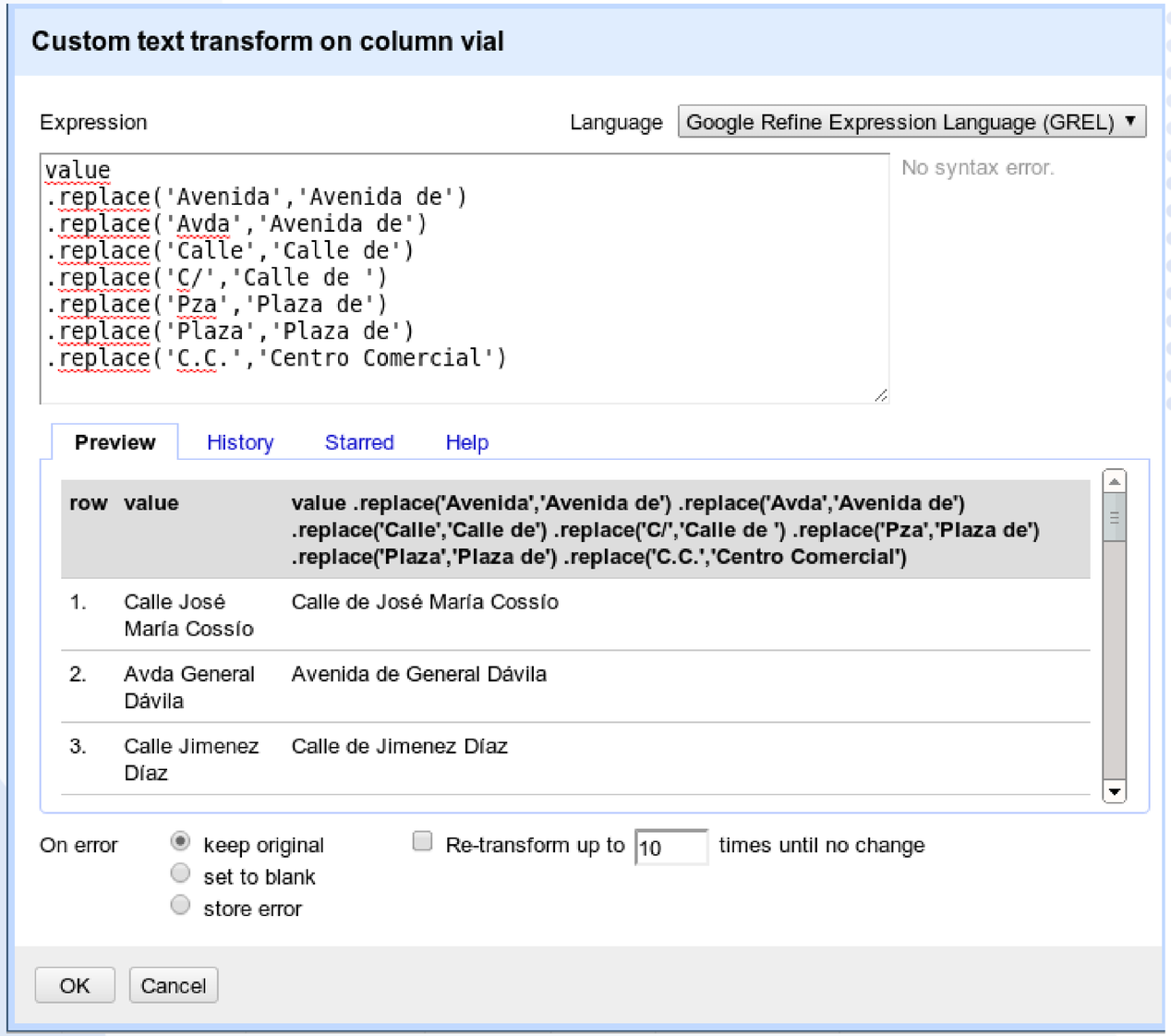
Por error
- (•) mantener original
- ( ) establecer en blanco
- ( ) error de tienda

[ ] Vuelva a transformar hasta 10 veces hasta que no haya cambios

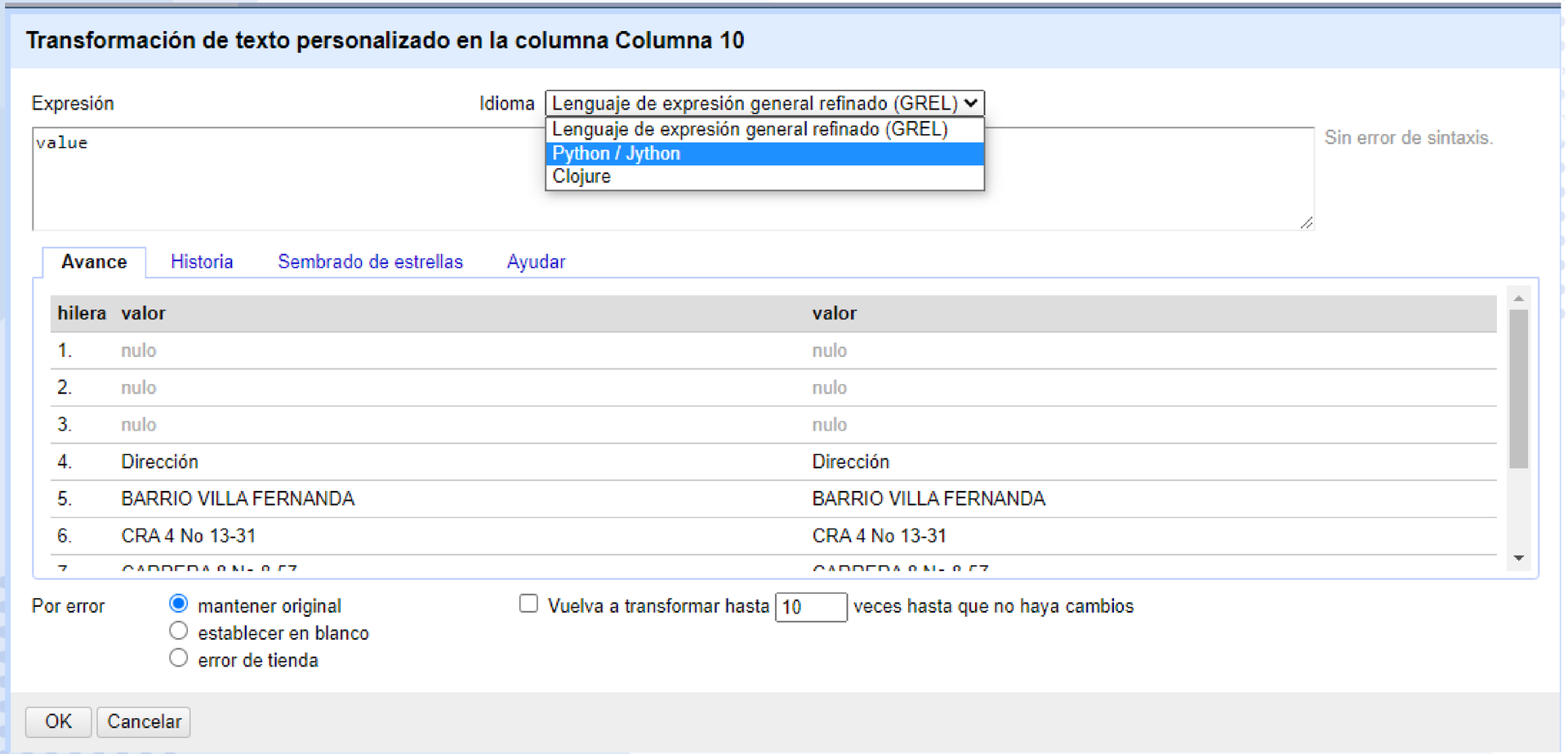
OK | Cancelar

El futuro digital es de todos | MinTIC

Las expresiones las podemos encadenar complicándolas tanto como queramos.
El ejemplo de la imagen siguiente realiza múltiples reemplazos de una sola vez.



El futuro digital es de todos | MinTIC



Es una herramienta de código abierto

Kit de herramientas que pueden encontrar https://gobiernodigital.mintic.gov.co/portal/Iniciativas/Software-libre/

MySQL Workbench

- Instalación de MySQL Workbench.
- Entorno y funcionalidades de MySQL Workbench.
- Ejercicio práctico en MySQL Workbench.

Open Refine

- Descargar Open Refine.
- Instalación Open Refine.
- Funcionalidades Open Refine.
- Datos Georreferenciados con Open Refine

Python:

Instalación de Python con pandas y Jupyter:
Comandos e Instrucciones Video1.txt.

- Funcionalidades y conceptos básicos:
- Comandos e Instrucciones Video2.txt.
- Entrada y visualización de datos desde datos.gov.co:

Comandos e Instrucciones Video3.txt..

Tock Chatbot:

Instalación de Tock:
Tock_Video1_Código.txt.

- Funcionalidades de Tock - Parte 1.
- Funcionalidades de Tock - Parte 2.

X-ROAD:

Conferencia sobre la implementación de X-ROAD, herramienta de interoperabilidad.

- Integración de servicios web a X-Road.
- Video 1 Generalidades X-Road.
- Video 2 Arquitectura y tecnología X-Road.
- Video 3 Implementación X Road.

El futuro digital es de todos

MinTIC

Conclusiones

Hemos demostrado cómo resolver problemas recurrentes tales como duplicados e inconsistencias ortográficas de forma automatizada con la ayuda de Open Refine, todos los datos están sucios, pero tú puedes hacer algo para remediarlo

No dudes en experimentar con las funciones de limpieza, ya que estás realizando estos pasos en una copia de tu conjunto de datos, y Open Refine te permite rastrear todos tus pasos (y volver atrás) en el caso de que hayas cometido un error.

El futuro digital es de todos

MinTIC

En este momento daremos unos minutos para llenar el formulario de diagnóstico

El futuro digital es de todos

MinTIC

PREGUNTAS