

Contenido

| | | |
|----|------------------------------------|----|
| 1. | Introducción | 2 |
| 2. | OntoRefine | 2 |
| 3. | Datos | 3 |
| 4. | Crear proyecto..... | 4 |
| 5. | Abrir proyecto | 5 |
| 6. | Cambios sobre filas/columnas | 6 |
| 7. | Crear RDF..... | 10 |
| a. | Esquema General | 10 |
| b. | Esquema modelado RDF | 11 |
| c. | Modelar RDF..... | 12 |
| d. | Insertar RDF..... | 13 |
| 8. | Resultados | 14 |

1. Introducción

GraphDB es una triple store, es decir, permite almacenar RDF. OntoRefine de GraphDB, a su vez, es una herramienta desarrollada para permitir a sus usuarios crear RDF a partir de datos como TSV, CSV, *SV, Excel (.xls y .xlsx), JSON, XML, RDF/XML, y Google Data documents.

Usando esta herramienta y lenguajes como SPARQL¹, GREL² y SPIN³, se puede formatear un archivo inicial en el formato elegido para modificar los datos, darles la estructura deseada, pasarlos a RDF, y subirlos a una triple store, concretamente, a GraphDB. Una vez almacenados los datos, se podrá trabajar con ellos, modificarlos y borrarlos, así como añadir nuevos.

A diferencia de otros medios de generar RDF como OpenRefine o Grafter, GraphDB permite insertar los datos generados directamente en la triple store, a diferencia de los mencionados anteriormente, en los que a partir del código generado había que valerse de otras herramientas o subir de forma manual el RDF generado a una triple store. Además, al permitir el uso de SPARQL, GREL, y SPIN no se necesitan definir funciones para realizar cambios sobre los datos.

Este manual esclarecerá mediante un ejemplo concreto las posibles dudas que puedan surgir al utilizar la nueva herramienta de GraphDB, OntoRefine. También explicará su utilización y ventajas respecto a otras formas de generar RDF desde datos tabulares.

2. OntoRefine

GraphDB OntoRefine es una herramienta de transformación de datos, basado en OpenRefine e integrado en el GraphDB Workbench. Se puede utilizar para convertir los datos tabulares en RDF e importarlo en un repositorio GraphDB, usando consultas SPARQL y un endpoint virtual.

OntoRefine permite:

- a. Subir un archivo de datos y crear un proyecto
- b. Ver los datos como RDF
- c. Transformar los datos usando diferentes funciones
- d. Importar el RDF creado directamente a un repositorio GraphDB usando el GraphDB SPARQL Endpoint

¹ SPARQL: <https://www.w3.org/TR/sparql11-query/>

² GREL: <https://github.com/OpenRefine/OpenRefine/wiki/GREL-Functions>

³ SPIN: <https://www.w3.org/Submission/spin-overview/>

3. Datos

Los datos utilizados son los correspondientes a la medición del aire en la ciudad de Vitoria-Gasteiz en una estación concreta, en este caso en av. Vitoria-Gasteiz. Entre ellos podemos encontrar además de la fecha en la que se produjo la medición, medidas de Benceno, CO, Etilbenceno, etc.

RDF tiene varios tipos de datos definidos: literales, tipos de datos numéricos, fechas, etc. Estos se utilizan para darle más información a quién lo esté procesando, por lo que a los datos que hay en el CSV de este ejemplo, dependiendo del tipo, interesará aplicarles distintas funciones para conseguir la estructura adecuada para que sean interpretados correctamente por el agente/humano correspondiente. A los datos numéricos, por ejemplo, se les debería aplicar una función para que sean identificados por datos `xsd:Integer`⁴ o `xsd:Double`, a las fechas, `xsd:Date` o `xsd:DateTime`.

En la siguiente figura se pueden ver gran parte de los datos:

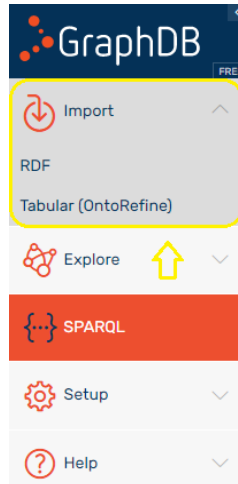
| Date | Benceno (µg/m3) | CO (mg/m3) | CO 8h (mg/m3) | CO 8h - Air Quality | Etilbenceno | NO (µg/m3) | NO2 (µg/m3) | NO2 - Air Quality | NOX (µg/m3) | Ortoxileno (µg/m3) | PM10 (µg/m3) | PM10 - Air Quality |
|------------|-----------------|------------|--------------------------|---------------------|-------------|------------|-------------|---------------------|-------------|--------------------|--------------|------------------------|
| 01/10/2016 | 0,11 | | 0,19 | Muy buena / Oso ona | 0,04 | 4 | 26 | Muy buena / Oso ona | 20 | | | 13 Muy buena / Oso ona |
| 30/09/2016 | 0,71 | | 0,41 | Muy buena / Oso ona | 1,1 | | 73 | Buena / Osa | 63 | 4,62 | 29 | Buena / Osa |
| 29/09/2016 | 0,32 | | Sin datos / Daturik gabe | | 0,72 | | 84 | Buena / Osa | 42 | 2,46 | 20 | Muy buena / Oso ona |
| 28/09/2016 | 0,31 | | 0,23 | Muy buena / Oso ona | 0,19 | 8 | 73 | Buena / Osa | 35 | | 7 | Muy buena / Oso ona |
| 27/09/2016 | 0,57 | | Sin datos / Daturik gabe | | 0,22 | 10 | 51 | Buena / Osa | 41 | 0,9 | 13 | Muy buena / Oso ona |
| 26/09/2016 | 0,3 | | 0,24 | Muy buena / Oso ona | 0,39 | 10 | 43 | Muy buena / Oso ona | 41 | 1,61 | 19 | Muy buena / Oso ona |
| 25/09/2016 | 0,21 | | 0,31 | Muy buena / Oso ona | 0,06 | | 38 | Muy buena / Oso ona | 16 | | 19 | Muy buena / Oso ona |
| 24/09/2016 | 0,47 | | 0,31 | Muy buena / Oso ona | 0,48 | 8 | 92 | Buena / Osa | 47 | 1,81 | 22 | Muy buena / Oso ona |
| 23/09/2016 | 0,38 | | 0,29 | Muy buena / Oso ona | 0,31 | 8 | 68 | Buena / Osa | 42 | | 19 | Muy buena / Oso ona |
| 22/09/2016 | 0,29 | | 0,23 | Muy buena / Oso ona | 0,17 | | 53 | Buena / Osa | 32 | 0,41 | 20 | Muy buena / Oso ona |
| 21/09/2016 | 0,39 | | 0,24 | Muy buena / Oso ona | 0,58 | 11 | 73 | Buena / Osa | 43 | 2,17 | 17 | Muy buena / Oso ona |
| 20/09/2016 | 0,2 | | 0,24 | Muy buena / Oso ona | 0,18 | | 42 | Muy buena / Oso ona | 28 | 0,59 | 11 | Muy buena / Oso ona |
| 19/09/2016 | 0,22 | | 0,28 | Muy buena / Oso ona | 0,24 | 17 | 51 | Buena / Osa | 46 | 0,81 | 13 | Muy buena / Oso ona |
| 18/09/2016 | | | 0,17 | Muy buena / Oso ona | 0,11 | | 35 | Muy buena / Oso ona | 15 | | 10 | Muy buena / Oso ona |
| 17/09/2016 | | | Sin datos / Daturik gabe | | 0,04 | | 24 | Muy buena / Oso ona | 13 | | 11 | Muy buena / Oso ona |
| 16/09/2016 | | | Sin datos / Daturik gabe | | 0,24 | | 47 | Muy buena / Oso ona | 17 | | 7 | Muy buena / Oso ona |
| 15/09/2016 | | | Sin datos / Daturik gabe | | 0,05 | | 11 | Muy buena / Oso ona | 8 | | 6 | Muy buena / Oso ona |
| 14/09/2016 | | | Sin datos / Daturik gabe | | 0,16 | | 41 | Muy buena / Oso ona | 24 | | 8 | Muy buena / Oso ona |
| 13/09/2016 | | | Sin datos / Daturik gabe | | 0,05 | | 30 | Muy buena / Oso ona | 15 | | 19 | Muy buena / Oso ona |
| 12/09/2016 | 0,38 | | 0,35 | Muy buena / Oso ona | 0,38 | | 70 | Buena / Osa | 34 | 1,48 | 30 | Buena / Osa |
| 11/09/2016 | 0,2 | | 0,21 | Muy buena / Oso ona | 0,08 | | 31 | Muy buena / Oso ona | 21 | | 18 | Muy buena / Oso ona |
| 10/09/2016 | 0,38 | | 0,19 | Muy buena / Oso ona | 0,35 | 7 | 58 | Buena / Osa | 35 | 1,36 | 21 | Muy buena / Oso ona |
| 09/09/2016 | 0,18 | | 0,19 | Muy buena / Oso ona | 0,1 | 6 | 31 | Muy buena / Oso ona | 27 | | 19 | Muy buena / Oso ona |
| 08/09/2016 | 0,15 | | 0,22 | Muy buena / Oso ona | 0,05 | 5 | 36 | Muy buena / Oso ona | 24 | | 14 | Muy buena / Oso ona |
| 07/09/2016 | 0,56 | | 0,33 | Muy buena / Oso ona | 0,55 | 15 | 82 | Buena / Osa | 53 | 2,24 | 41 | Buena / Osa |
| 06/09/2016 | 0,41 | | 0,29 | Muy buena / Oso ona | 0,32 | 16 | 70 | Buena / Osa | 64 | 0,85 | 28 | Buena / Osa |
| 05/09/2016 | 0,69 | | 0,31 | Muy buena / Oso ona | 0,22 | 16 | 72 | Buena / Osa | 52 | 0,79 | 34 | Buena / Osa |
| 04/09/2016 | 0,35 | | 0,28 | Muy buena / Oso ona | 0,25 | 3 | 45 | Muy buena / Oso ona | 31 | 0,64 | 31 | Buena / Osa |
| 03/09/2016 | 0,19 | | 0,27 | Muy buena / Oso ona | 0,22 | | 38 | Muy buena / Oso ona | 27 | 0,51 | 23 | Muy buena / Oso ona |
| 02/09/2016 | 0,38 | | 0,3 | Muy buena / Oso ona | 0,21 | | 60 | Buena / Osa | 45 | | 27 | Buena / Osa |
| 01/09/2016 | 0,23 | | 0,26 | Muy buena / Oso ona | 0,18 | | 60 | Buena / Osa | 41 | | 20 | Muy buena / Oso ona |

⁴ Datatypes RDF: <https://www.w3.org/TR/swbp-xsch-datatypes/>

4. Crear proyecto

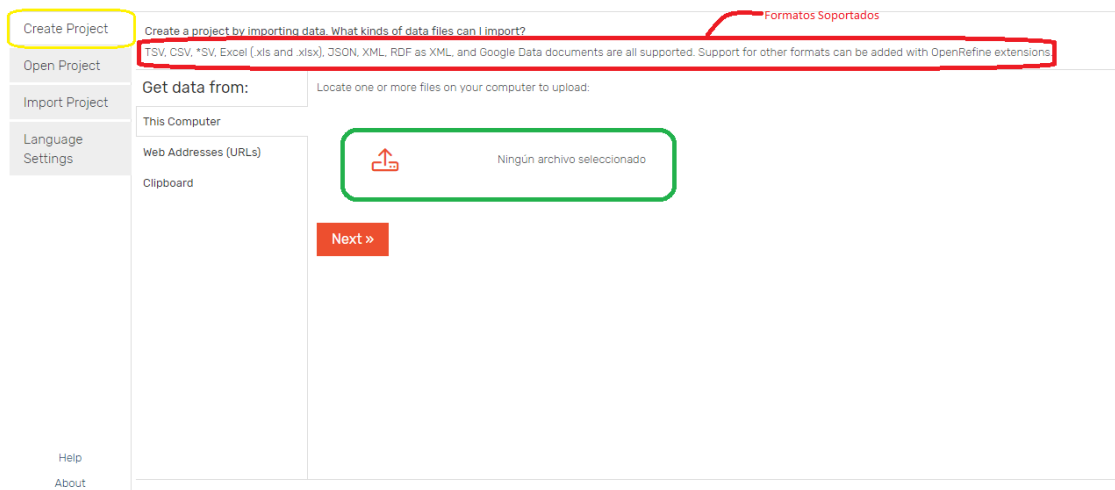
El repositorio⁵ de GraphDB usado en este ejemplo fue el propio de Eurohelp.

- a. Dentro de un repositorio, click en Import -> Tabular (OntoRefine).




- b. Se selecciona el archivo a subir.

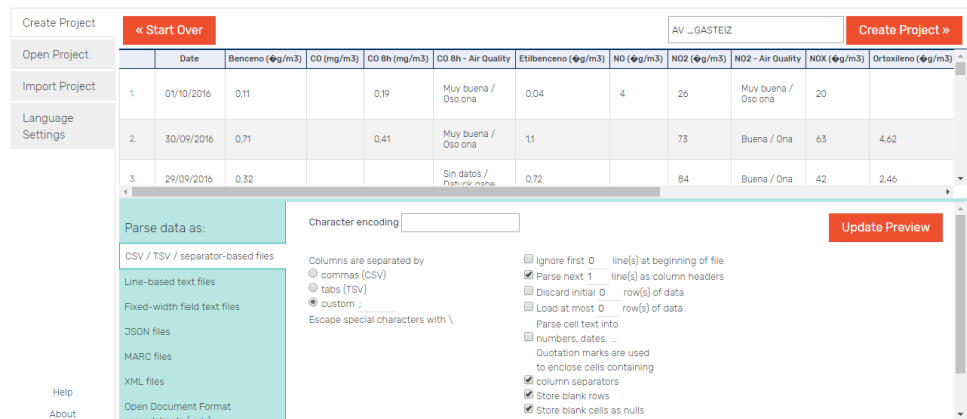
OntoRefine ?



- c. Next.

⁵ Repositorio de GraphDB de Eurohelp: <http://lod.eurohelp.es:7200/>

- d. (Opcional) Cambiar nombre al proyecto, y/o las configuraciones de la tabla.
OntoRefine 



Create Project

« Start Over

AV _GASTEIZ

Create Project »

| | Date | Benceno (µg/m3) | CO (mg/m3) | CO 8h (mg/m3) | CO 8h - Air Quality | Etilbenceno (µg/m3) | NO (µg/m3) | NO2 (µg/m3) | NO2 - Air Quality | NOX (µg/m3) | Ortolano (µg/m3) |
|----|------------|-----------------|------------|---------------|----------------------------|---------------------|------------|-------------|---------------------|-------------|------------------|
| 1. | 01/10/2016 | 0.11 | | 0.19 | Muy buena / Oso ona | 0.04 | 4 | 26 | Muy buena / Oso ona | 20 | |
| 2. | 30/09/2016 | 0.71 | | 0.41 | Muy buena / Oso ona | 1.1 | | 73 | Buena / Ona | 63 | 4.62 |
| 3. | 29/09/2016 | 0.32 | | | Sin datos / Sin info. onna | 0.72 | | 84 | Buena / Ona | 42 | 2.46 |

Parse data as:

Character encoding

Update Preview

CSV / TSV / separator-based files

Line-based text files

Fixed-width field text files

JSON files

MARC files

XML files

Open Document Format

Columns are separated by

Ignore first 0 line(s) at beginning of file

Parse next 1 line(s) as column headers

Discard initial 0 row(s) of data

Load at most 0 row(s) of data

Parse cell text into numbers, dates, ...

Quotation marks are used to enclose cells containing column separators

Store blank rows

Store blank cells as nulls

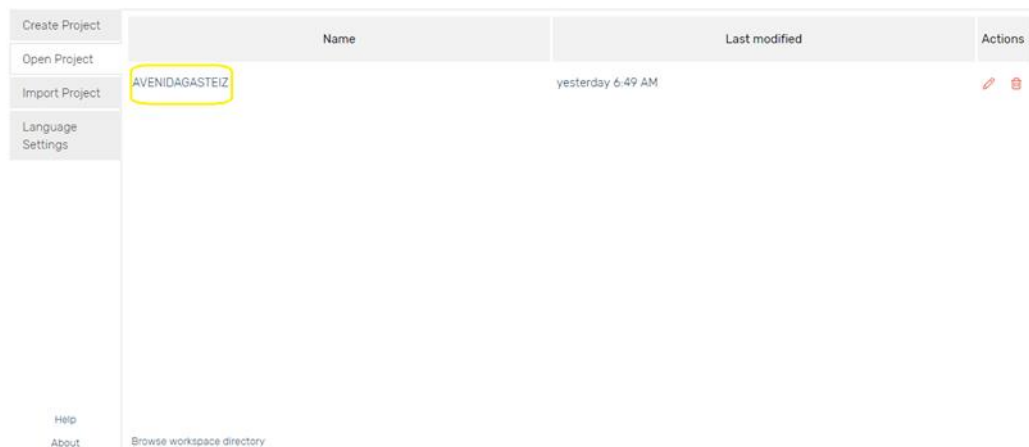
- e. Click Create Project.

5. Abrir proyecto

Una vez se haya creado el proyecto:

- a. Click Open Project.

OntoRefine 





Create Project

Open Project

Import Project

Language Settings

| Name | Last modified | Actions |
|----------------|-------------------|---|
| AVENIDAGASTEIZ | yesterday 6:49 AM |   |

Help

About

Browse workspace directory


- b. Seleccionar el proyecto que se desea abrir.

6. Cambios sobre filas/columnas

Una vez creado o abierto un proyecto se genera una tabla sobre la cual se pueden realizar modificaciones sobre las columnas/filas existentes. OntoRefine da la posibilidad de forma similar a OpenRefine de realizar cambios manualmente sin necesidad de codificar instrucciones, y/o haciendo click en SPARQL empezar a modelar RDF. Los cambios sobre filas/columnas que se mencionarán a continuación se pueden realizar también utilizando funciones SPARQL y GREL, es decir, todo lo realizado manualmente se puede hacer mediante código.

A pesar de que todos los cambios deseados sobre un CSV se puedan realizar utilizando funciones, en ocasiones, y dependiendo del tipo de cambio que se va a realizar resulta más adecuado realizarlo de forma manual, ya que si se codifica puede ser necesaria una cantidad significativa de código y si ese cambio se va a realizar sobre varias columnas, ese código ya significativo, se incrementaría.

OntoRefine

name: calidadDelAireAvenidaGasteiz 

Open... Export Help

Modelar RDF SPARQL

Facet / Filter Undo / Redo 0

275 rows

Show as: **rows** records Show: 5 10 25 **50** rows

« first < previous 1 - 50 next > last »

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started? Watch these screencasts

Cambios en filas/columnas

| | | date | Benceno (µg/l) | CO (mg/m3) | CO 8h (mg/m3) | CO 8h - Air Que | Etilbenceno (µg/l) | NO (µg/m3) | NO2 (µg/m3) | N02 |
|---|-----|------------|----------------|------------|---------------|--------------------------|--------------------|------------|-------------|-------|
| ☆ | 1. | 01/10/2016 | 0.11 | | 0.19 | Muy buena / Oso | 0.04 | 4 | 26 | Muy t |
| ☆ | 2. | 30/09/2016 | 0.71 | | 0.41 | Muy buena / Oso | 1.1 | | 73 | Buen |
| ☆ | 3. | 29/09/2016 | 0.32 | | | Sin datos / Daturik gabe | 0.72 | | 84 | Buen |
| ☆ | 4. | 28/09/2016 | 0.31 | | 0.23 | Muy buena / Oso | 0.19 | 8 | 73 | Buen |
| ☆ | 5. | 27/09/2016 | 0.57 | | | Sin datos / Daturik gabe | 0.22 | 10 | 51 | Buen |
| ☆ | 6. | 26/09/2016 | 0.3 | | 0.24 | Muy buena / Oso | 0.39 | 10 | 43 | Muy t |
| ☆ | 7. | 25/09/2016 | 0.21 | | 0.31 | Muy buena / Oso | 0.06 | | 38 | Muy t |
| ☆ | 8. | 24/09/2016 | 0.47 | | 0.31 | Muy buena / Oso | 0.48 | 8 | 92 | Buen |
| ☆ | 9. | 23/09/2016 | 0.38 | | 0.29 | Muy buena / Oso | 0.31 | 8 | 68 | Buen |
| ☆ | 10. | 22/09/2016 | 0.29 | | 0.23 | Muy buena / Oso | 0.17 | | 53 | Buen |
| ☆ | 11. | 21/09/2016 | 0.39 | | 0.24 | Muy buena / Oso | 0.58 | 11 | 73 | Buen |
| ☆ | 12. | 20/09/2016 | 0.2 | | 0.24 | Muy buena / Oso | 0.18 | | 42 | Muy t |
| ☆ | 13. | 19/09/2016 | 0.22 | | 0.28 | Muy buena / Oso | 0.24 | 17 | 51 | Buen |

1. **Cambiar nombre columna:** Tiene bastante utilidad, dado que cuando hay nombres de columnas con simbología, ya sean acentos, símbolos o en menor medida espacios, OntoRefine no permite recuperar los valores de las columnas para realizar cambios sobre ellos. En este caso, los datos tenían acentos y espacios, además de simbología para la unidad de medición, por lo que se les cambió el nombre a las columnas para poder trabajar con ellas.

- a. En la columna deseada: Edit column->Rename this column/Remove column.

OntoRefine

name: calidadDelAireAvenidaGasteiz 

Open... Export Help

Facet / Filter Undo / Redo 0

275 rows

Show as: rows records Show: 5 10 25 50 rows

« first < previous 1 - 50 next > last »

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started? Watch these screencasts

| | Date | Benceno (µg/m³) | CO (mg/m³) | CO 8h (mg/m³) | CO 8h - Air Qual | Etibenceno (µg/m³) | NO (µg/m³) | NO2 (µg/m³) | NO2 |
|-----|------------|-----------------|------------------------------------|---------------|------------------|--------------------|------------|-------------|----------|
| 1. | 01/10/2016 | Facet > | | 0,19 | Muy buena / Oso | 0,04 | 4 | 26 | Muy t... |
| 2. | 30/09/2016 | Text filter | | 0,41 | Muy buena / Oso | 1,1 | | 73 | Buen... |
| 3. | 29/09/2016 | Edit cells > | | | Sin datos / | 0,72 | | 64 | Buen... |
| 4. | 28/09/2016 | Edit column > | Split into several columns... | | ina / Oso | 0,19 | 8 | 73 | Buen... |
| 5. | 27/09/2016 | Transpose > | Add column based on this column... | | s / | 0,22 | 10 | 51 | Buen... |
| 6. | 26/09/2016 | Sort... | Add column by fetching URLs... | | ina / Oso | 0,39 | 10 | 43 | Muy t... |
| 7. | 25/09/2016 | View > | Rename this column | | na / Oso | 0,06 | | 38 | Muy t... |
| 8. | 24/09/2016 | Reconcile > | Remove this column | | ina / Oso | 0,48 | 8 | 92 | Buen... |
| 9. | 23/09/2016 | | Move column to beginning | | ina / Oso | 0,31 | 8 | 68 | Buen... |
| 10. | 22/09/2016 | | Move column to end | | ina / Oso | 0,17 | | 53 | Buen... |
| 11. | 21/09/2016 | | Move column left | | ina / Oso | 0,58 | 11 | 73 | Buen... |
| 12. | 20/09/2016 | | Move column right | | | | | | |
| 13. | 19/09/2016 | | | | | | | | |

2. **Convertir campos con valores numéricos enteros a formato xsd:Integer:** Se debe realizar porque de otra forma, los números serían tipificados automáticamente como literales en lugar de números en formato RDF.
 - a. En la columna deseada: Edit cells ->Common transforms-> To number.

OntoRefine

name: calidadDelAireAvenidaGasteiz 

Open... Export Help

Facet / Filter Undo / Redo 0

275 rows

Show as: rows records Show: 5 10 25 50 rows

« first < previous 1 - 50 next > last »

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started? Watch these screencasts

| | Air Qual | Etibenceno (µg/m³) | NO (µg/m³) | NO2 (µg/m³) | NO2 - Air Qual | NOX (µg/m³) | Ortoxileno (µg/m³) | PM10 (µg/m³) | PM10 - Air Qual | PM2.5 |
|-----|----------|--------------------|-----------------------------|--------------------------------------|-----------------|-------------|--------------------|--------------|-----------------|-------|
| Oso | 0,04 | Facet > | | | Muy buena / Oso | 20 | | 13 | Muy buena / Oso | 6 |
| Oso | 1,1 | Text filter | | | Buena / Oso | 63 | 4,62 | 29 | Buena / Oso | 12 |
| Oso | 0,72 | Edit cells > | Transform... | | | 42 | 2,46 | 20 | Muy buena / Oso | 8 |
| Oso | 0,19 | Edit column > | Common transforms > | Trim leading and trailing whitespace | | | | 7 | Muy buena / Oso | 4 |
| Oso | 0,22 | Transpose > | Fill down | Collapse consecutive whitespace | | | | 13 | Muy buena / Oso | 7 |
| Oso | 0,39 | Sort... | Blank down | Unescape HTML entities | | | | 19 | Muy buena / Oso | 6 |
| Oso | 0,06 | View > | Split multi-valued cells... | To titlecase | | | | 19 | Muy buena / Oso | 6 |
| Oso | 0,48 | Reconcile > | Join multi-valued cells... | To uppercase | | | | 22 | Muy buena / Oso | 9 |
| Oso | 0,31 | | Cluster and edit... | To lowercase | | | | 19 | Muy buena / Oso | 8 |
| Oso | 0,17 | | | To number | | | | 20 | Muy buena / Oso | 10 |
| Oso | 0,58 | | | To date | | | | 17 | Muy buena / Oso | 10 |
| Oso | 0,18 | | | To text | | | | 11 | Muy buena / Oso | 7 |
| Oso | 0,24 | | | Blank out cells | | | | 13 | Muy buena / Oso | 7 |

3. **Convertir campos con valores decimales a formato xsd:Double:** Como en el caso anterior, si no se les aplicara el formato adecuado, este tipo de datos serían tomados como literales, por lo cual hay que tipificarlos adecuadamente.
 - a. En la columna deseada: Edit cells ->Transform....

Open... Export Help

Not sure how to get started?
Watch these screencasts

- Cancel

- 8

name: calidadDelAireAvenidaGasteiz

name: calidadDelAireAvenidaGasteiz ↗

Open... Export Help

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
Watch these screencasts:

- a. En la columna deseada: Edit column->Split into several colums.

name: calidadDelAireAvenidaGasteiz

name: calidadDelAireAvenidaGasteiz

Open... Export Help

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
Watch these screencasts

- 9

Split column CO 8h - Air Quality into several columns

How to Split Column

☒ by separator

Separator ☐ regular expression

Split into columns at most (leave blank for no limit)

☐ by field lengths

List of integers separated by commas, e.g., 5, 7, 15

After Splitting

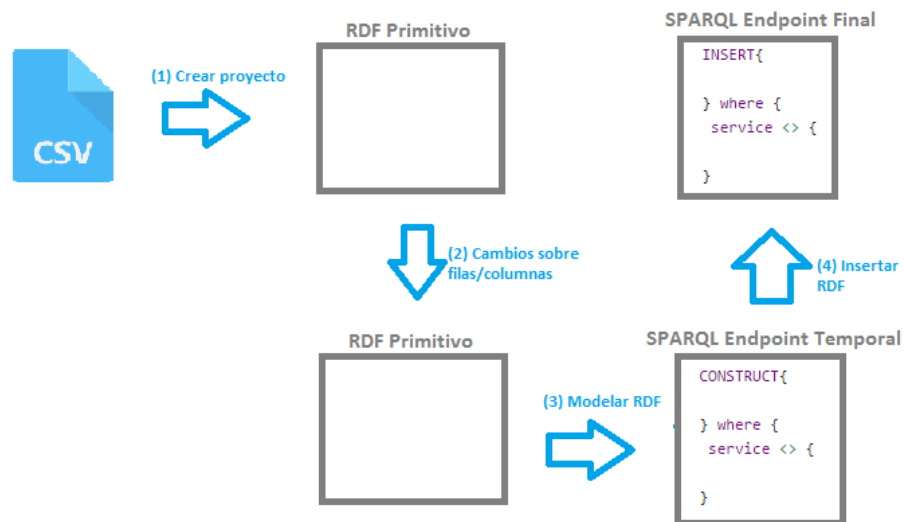
☒ Guess cell type

☒ Remove this column

OK Cancel

7. Crear RDF

a. Esquema general



1. **Crear proyecto:** Se refiere tanto a la creación de un proyecto en OntoRefine a partir de un CSV, como a la carga de un proyecto ya creado. OntoRefine crea, a partir del CSV que se le ha pasado un RDF primitivo que representa los datos y columnas de éste, el RDF primitivo se usará para que el usuario pueda referirse a esos datos cuando esté modelando el RDF.
2. **Cambios en filas /columnas:** Sirve para aplicar de forma manual cambios sobre los datos existentes. En este paso todavía no se está generando aún un nuevo RDF, sólo se están realizando alteraciones sobre el RDF primitivo creado por OntoRefine. Como se mencionará más adelante los cambios realizados en esta parte se pueden realizar también cuando se esté modelando el RDF.
3. **Modelar RDF:** Se refiere al proceso de creación del RDF, al uso de ontologías y estándares para darle la estructura deseada. A lo largo del documento se aclarará este punto con ejemplos. Mientras se esté construyendo el RDF, desde el SPARQL Endpoint Temporal se debe hacer uso de la cláusula construct, que mapea los datos RDF sin procesar y devuelve un grafo, lo que permite

realizar un seguimiento del proceso siendo posible la visualización del grafo siempre que se desee.

4. **Insertar RDF:** Una vez concluido el proceso de modelado del RDF, para insertar los datos en el repositorio se debe cambiar la cláusula construct por insert, que en lugar de mostrar el RDF, lo insertará en el repositorio en el que se encuentre. Esto se hará sobre el SPARQL Endpoint Final que haciendo uso de service, se conectará con el SPARQL Endpoint Temporal.

b. Esquema modelado RDF

```

PREFIX pr: <http://purl.org/ontology/prv/core#>
PREFIX qb: <http://purl.org/linked-data/cube#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix spif: <http://spinrdf.org/spif#>
prefix sdmx-attribute: <http://purl.org/linked-data/sdmx/2009/attribute#>
prefix sdmx-measure: <http://purl.org/linked-data/sdmx/2009/measure#>
prefix medicion: <http://opendata.euskadi.eus/medicion/AV-GASTEIZ/>
prefix estacion: <http://opendata.euskadi.eus/estacion/>

INSERT {
  GRAPH <http://opendata.euskadi.eus/graph/AirQuality> {
    ?uriICAEstacion a qb:Observation;
    rdfs:comment ?ICAEstacionAQComment;
    dct:date ?fechaParseada;
    geo:location estacion:AV-GASTEIZ;
    sdmx-measure:obsValue ?ICAEstacionEsp;
    sdmx-measure:obsValue ?ICAEstacionEusk.
  }
} where {
  service <http://lod.eurohelp.es:7200/rdf-bridge/2326104478520> {
    ?avGasteizRow a <urn:Row> ;
    <urn:col:Date> ?date ;
    <urn:col:ICAEstacion1> ?ICAEstacion1;
    <urn:col:ICAEstacion2> ?ICAEstacion2.

    bind(spif:parseDate(?date, "dd/MM/yyyy") as ?fechaParseada)
    #URIS elementos
    bind(iri(concat("http://opendata.euskadi.eus/estacion/AV-GASTEIZ/", spif:encodeURL(?fechaParseada), "/ICAEstacion" )) as ?uriICAEstacion)

    bind(STRLANG(?ICAEstacion1,"es") as ?ICAEstacionEsp)
    bind(STRLANG(?ICAEstacion2,"eu") as ?ICAEstacionEusk)

    bind(STRLANG("The value of Benceno in a determinate date","en") as ?bencenoComment)
    bind(STRLANG("ICA Estacion Air Quality in a determinate date","en") as ?ICAEstacionAQComment)

    bind(iri("http://dd.eionet.europa.eu/vocabulary/uom/concentration/ug.m-3") as ?umcubico)
  }
}

```

1 Prefix que se usarán para la creación del RDF

2 Creación de los nodos del grafo

3 Obtenemos de la tabla las variables que vamos a utilizar

4 Aplicamos funciones a las variables obtenidas con anterioridad

1. Se refiere a los estándares u ontologías que se necesitan para modelar el RDF para darle la estructura deseada. En nuestro caso, por ejemplo, vamos a crear un RDF que son mediciones y lo haremos diciendo que los datos son qb:Observation según lo que define el estándar Data Cube.
2. Esta parte se refiere a la estructura que tendrán los nodos del grafo y podrá ser encapsulada dentro de insert o construct. Para nuestras mediciones, además de fijar que van a ser un qb:Observation de Data Cube creamos más triples para añadir comentarios sobre una medición concreta, su localización, la unidad en la que se ha tomado la medida y el valor de ésta. Por ejemplo, al poner ?uriBenceno rdfs:comment ?bencenoComment se está añadiendo un comentario al recurso ?uriBenceno.
3. Recogemos los datos deseados de la tabla por medio de <urn:col:nombreColumna>, y los asignamos a variables para su posterior utilización. Esto se puede llevar a cabo porque al cargar el CSV en OntoRefine este crea un RDF primitivo que contiene ya la información inicial

proporcionada por el CSV. Al usar la anterior instrucción nos estamos refiriendo a ese RDF.

4. Sobre las variables anteriormente recogidas se pueden aplicar funciones de SPARQL y SPIN. Dado que OntoRefine recoge los datos de la tabla tomándolos a todos como literales este paso tiene suma importancia porque sin él los datos no tendrían diferencias entre sí, ni la estructura adecuada. En el ejemplo se le da formato a la fecha para fijar que es tipo `xsd:DateTime` de RDF, se aplica funciones para fijar que un literal es un recurso, etc.

c. Modelar RDF

Tras hacer click en SPARQL, se cargará una ventana como la siguiente donde se puede empezar a escribir código SPARQL.



1. **Conversión literales a recursos:** Como se mencionaba anteriormente, OntoRefine “entiende” todos los datos como literales, por lo que para usar una URI y que este la “entienda” como tal hay que aplicarles una función. Se utiliza la función `iri(literal URI)` que convierte lo que se le pasa como parámetro a un recurso, y la función `bind (expression AS ?var)`, que asigna un valor a una determinada variable.

```
bind(iri("http://dd.eionet.europa.eu/vocabulary/uom/concentrati  
on/ug.m-3") as ?umcubico)
```

2. **Conversión de literales a `xsd:Date`:** Para convertir literales a formato `xsd:Date` de RDF se hace uso de la función `spif:parseDate(literal date, literal format)`

```
bind(spif:parseDate(?date, "dd/MM/yyyy") as ?fechaParseada)
```

3. **Poner etiqueta de idioma:** Al poner literales, es recomendable fijar el idioma en el que están. Como se comentaba antes en los datos tenemos valores de mediciones en varios idiomas, por lo que conviene especificar en qué idioma está ese dato. Para llevarlo a cabo se usa la función `STRLANG(literal lexicalForm, literal langTag)`

```
bind(STRLANG (?ICAEstacion2, "eu") as ?ICAEstacionEusk)
```

d. Insertar RDF

1. Una vez modelado el RDF, click en SPARQL Endpoint.



2. Copiar y pegar la URI, debajo de where, usando service.

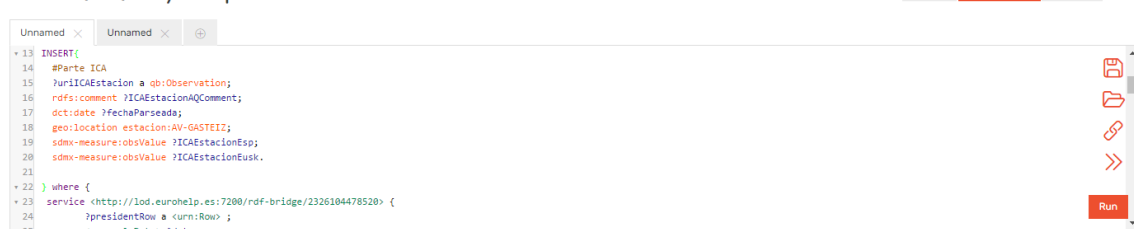


3. Hacer click en SPARQL.



4. Pegar el código creado anteriormente, y cambiar construct por insert, para que en vez de mostrarnos el grafo, lo inserte en el repositorio.

SPARQL Query & Update



The screenshot shows a web interface for SPARQL queries. At the top, there are tabs for 'Editor only', 'Editor and results', and 'Results only'. Below the tabs, there is a text area containing a SPARQL INSERT query. The query is as follows:

```
13 INSERT {
14   #Parte ICA
15   ?uriICAEstacion a qb:Observation;
16   rdfs:comment ?ICAEstacionAQComment;
17   dct:date ?fechaParseada;
18   geo:location estacion:AV-GASTEIZ;
19   sdmx-measure:obsValue ?ICAEstacionEsp;
20   sdmx-measure:obsValue ?ICAEstacionEusk.
21 }
22 where {
23   service <http://lod.eurohelp.es:7200/rdf-bridge/2326104478520> {
24     ?presidentRow a <urn:Row> ;
```

5. Hacer click en Run.

8. Resultados

Los resultados obtenidos, la sentencia SPARQL que da lugar a los datos RDF y el CSV utilizado mencionado se pueden encontrar en: <https://github.com/mishel-uchuari/ConverterCSVtoRDF-SPARQL-OPENREFINE-GRAPHDB.git>

Los datos usados en este documento se recogieron de Open Data Euskadi, <http://opendata.euskadi.eus/catalogo/-/calidad-aire-en-euskadi-2016/> , concretamente del conjunto de datos: “Datos validados - Diarios y horarios”, y dentro de ellos, el CSV concreto, AV_GASTEIZ.