



UNIVERSIDAD DE MÁLAGA



Graduado en Ingeniería de la Salud

## Proyecto Estándares de datos abiertos

Open data standards project

Realizado por

Jesús Aldana Martín, Hugo Ávalos de Rorthais, Pablo Molina Sánchez  
y Juan Carlos Ruiz Ruiz

Tutorizado por

Laura Panizo Jaime

Departamento

Lenguajes y Ciencias de la Computación

MÁLAGA, diciembre de 2022



UNIVERSIDAD  
DE MÁLAGA



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA INFORMÁTICA  
GRADUADO EN INGENIERÍA DE LA SALUD

**Proyecto de estándares de datos abiertos**

**Open Data Standards project**

Realizado por  
**Jesús Aldana Martín, Hugo Ávalos de  
Rorthais,  
Pablo Molina Sánchez, Juan Carlos Ruiz Ruiz**

Tutorizado por  
**Laura Panizo Jaime**

Departamento  
**Lenguajes y Ciencias de la Computación**

UNIVERSIDAD DE MÁLAGA  
MÁLAGA, DICIEMBRE DE 2022

Fecha defensa: diciembre de 2022



# Índice

<b>1. Introducción</b>	<b>5</b>
1.1. Motivación . . . . .	5
1.2. Objetivos . . . . .	6
1.3. Estructura del documento . . . . .	6
1.4. Tecnologías usadas . . . . .	7
<b>2. XML</b>	<b>9</b>
2.1. Creación del XML . . . . .	9
2.2. Consultas XQuery . . . . .	11
2.2.1. Número de accidentes de mujeres . . . . .	11
2.2.2. Número de accidentes en cada día de la semana . . . . .	11
2.2.3. Tipo de accidentes según cada condición de la carretera . . . . .	13
2.3. Transformaciones XSLT . . . . .	15
2.3.1. De XML a XML . . . . .	15
2.3.2. De XML a HTML . . . . .	17
<b>3. Ontología</b>	<b>19</b>
3.1. Grafo RDF . . . . .	19



# 1

# Introducción

En la asignatura de *Estándares de datos abiertos e integración de datos abiertos* vamos a realizar un proyecto en el que generaremos una ontología de un conjunto de datos para poder mostrarla mediante transformaciones XSLT y gráficos de OWL.

## 1.1. Motivación

La idea del proyecto es utilizar un conjunto de datos en CSV para generar un XML utilizable y a partir de dicho XML obtendremos un conjunto de datos con el que podremos generar la ontología. Usaremos datos sobre los accidentes de bicicletas en Reino Unido, asociados al perfil del ciclista y al estado del tiempo en ese momento. Los datos los obtendremos del siguiente CSV:

[Bicycle accidents in Great Britain 1979 to 2018](#)

A partir de estos datos, obtendremos un único XML con el que podremos utilizar los datos y generar la base de datos.

Por otro lado buscamos generar una página web donde podamos mostrar todos los datos obtenidos y mostrar los grafos de la red correspondientes. Esta web será meramente para mostrar los resultados obtenidos y facilitar la visualización de los datos. La web tendrá una forma como la mostrada en la figura 4.

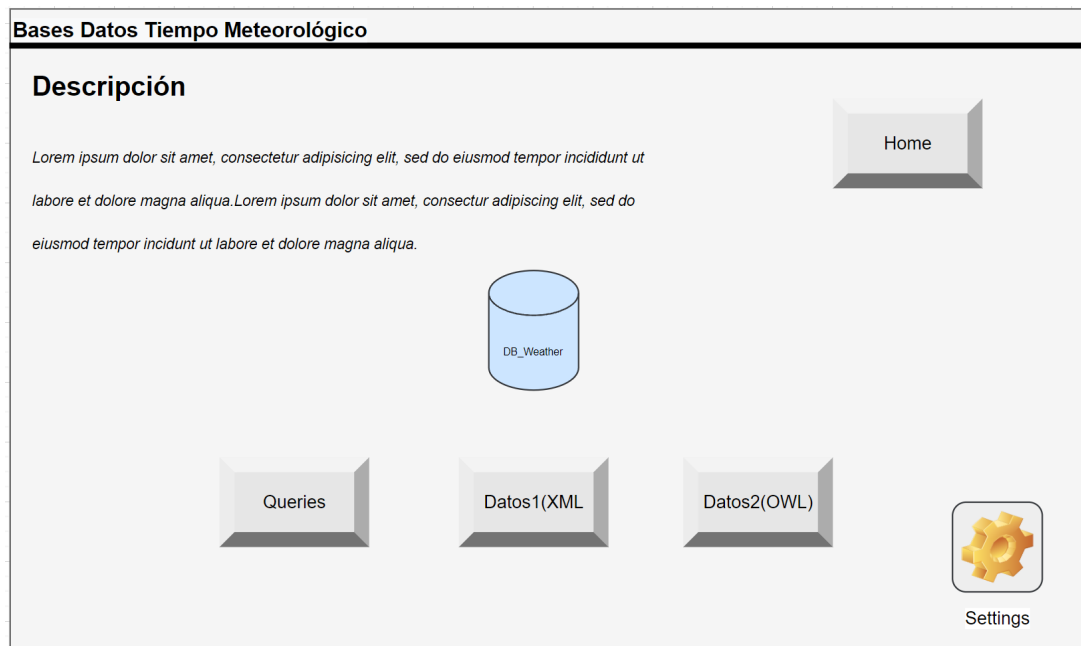


Figura 1: Diseño inicial de la página web

## 1.2. Objetivos

El objetivo principal del proyecto es aprender a utilizar las tecnologías de manejo de bases de datos XML, transformaciones XSLT y utilizar tecnologías de generación de ontologías en OWL y RDF.

Como objetivo secundario tenemos el generar una ontología que sea ampliable y explique de forma detallada la distribución de accidentes de bicicleta en diferentes condiciones y horarios, y obtener información que relacione los datos de forma automática.

## 1.3. Estructura del documento

El documento constará de dos partes diferenciadas, una primera parte en la que se tratarán los conceptos de XML, XSLT y Xquery; y una segunda parte en la que usaremos OWL, RDF y SPARQL.

En la primera parte explicaremos como hacemos la transformación de nuestros datos de CSV a XML, también realizaremos varias transformaciones XSLT de XML a HTML que mostraremos en nuestra página web, por último utilizaremos Xqueries para obtener datos concretos de nuestro XML y mostrarlos, de nuevo, en nuestra página web.

En la segunda parte, nos centraremos en tecnologías usadas para el manejo de ontologías como serán RDF, en primer lugar, para generar una primera ontología. Además generaremos la ontología también en OWL y mostraremos su grafo en nuestra página web. Por último, usaremos queries de SPARQL para manejar nuestros datos y mostrarlos en la web.

#### **1.4. Tecnologías usadas**

Para el desarrollo de la primera parte utilizaremos el Visual Studio Code, para realizar las transformaciones a XML, y el ExistDB, para las consultas Xquery. Para el desarrollo de la segunda parte utilizaremos Protegé, para la creación de la ontología, y el GraphDB para las consultas en SPARQL.

Todo código utilizado en el proyecto se podrá obtener del siguiente repositorio:

<https://github.com/JCruiz15/Estandares-project>





# 2

## XML

### 2.1. Creación del XML

Lo primero que hemos tenido que hacer ha sido generar un XML a partir de nuestros datos en CSV originales. Nuestro CSV contenía información acerca de todos los accidentes de bicicletas producidos en Reino Unido desde 1979 hasta 2018. Este dataset al ser excesivamente grande para nuestro cometido en el proyecto hemos decidido utilizar 107 elementos del CSV, comprendiendo aproximadamente los accidentes producidos durante 1995.

En primer lugar para generar el XML hemos creado un XML schema que nos ayudase a verificar la correcta implementación de los datos. Este schema nos ha permitido decidir como vamos a disponer los datos en el XML final.

Una vez teníamos el schema, utilizamos la web <https://www.convertcsv.com/csv-to-xml.htm> para obtener el XML resultante ya que el volumen de datos que habría que traducir a XML sería excesivamente grande. Gracias a esta web y a la siguiente plantilla obtenida a partir del schema obtendremos el XML:

```
<?xml version="1.0"?>{br}<accidents>{br}

<Accident {h1}="{f1}">
  <N_vehicules>{f2}</N_vehicules>
  <N_casualties>{f3}</N_casualties>
  <{h4}>{f4}</{h4}>
  <{h5}>{f5}</{h5}>
  <Week_day>{f9}</Week_day>
  <Conditions>
    <{h6}>{f6}</{h6}>
    <{h7}>{f7}</{h7}>
    <{h10}>{f10}</{h10}>
```

```

    <{h8}>{f8}</{h8}>
    <{h11}>{f11}</{h11}>
  </Conditions>
  <Biker>
    <{h12}>{f12}</{h12}>
    <{h14}>{f14}</{h14}>
    <{h13}>{f13}</{h13}>
  </Biker>
</Accident>

</accidents>

```

Así con la plantilla obtenemos el XML final que utilizaremos para realizar las Xqueries y las transformaciones. La forma del XML obtenido sería el siguiente (visualizamos la primera instancia).

```

<?xml version="1.0"?>
<?xml-model href="schema_bicycle_accidents.xsd" type="application/xml"
  schematypens="http://www.w3.org/2001/XMLSchema"?>
<accidents>
  <Accident Accident_Index="19954100B0030">
    <N_vehicules>2</N_vehicules>
    <N_casualties>1</N_casualties>
    <Date>1995-01-25</Date>
    <Time>22:45</Time>
    <Week_day>Wednesday</Week_day>
    <Conditions>
      <Speed_limit>30.0</Speed_limit>
      <Road_conditions>Wet</Road_conditions>
      <Road_type>Single carriageway</Road_type>
      <Weather_conditions>Clear and windy</Weather_conditions>
      <Light_conditions>Darkness lights lit</Light_conditions>
    </Conditions>
    <Biker>
      <biker_gender>Male</biker_gender>
      <biker_age_group>26 to 35</biker_age_group>
      <biker_severity_injury>Slight</biker_severity_injury>
    </Biker>
  </Accident>
</accidents>

```

```
</Biker>
</Accident>
...
</accidents>
```

---

## 2.2. Consultas XQuery

Lo siguiente que haremos será implementar tres consultas XQuery. Una consulta XQuery nos permite obtener información concreta a partir de nuestro XML. Hemos implementado una consulta para conocer cuantas mujeres han sufrido accidentes en bicicleta, otra que nos muestra información sobre cuantos accidentes se han producido en cada día de la semana y por último hemos creado una tercera que nos muestra las condiciones en las que puede encontrarse la carretera.

### 2.2.1. Número de accidentes de mujeres

En la primera Xquery que hemos hecho, hemos calculado el número de mujeres que han tenido un accidente en bici. Este ha sido el código utilizado:

```
xquery version "3.1";
count(doc('/db/apps/doc/xml_bycicle_accidents.xml')/accidents/Accident/
  Biker[biker_gender="Female"])
```

---

El resultado obtenido ha sido el siguiente:

---

18

---

Como podemos ver, es un número bastante bajo respecto al total (107), suponiendo el número de accidentes en bici protagonizado por mujeres solo un 16,8 por ciento del total.

### 2.2.2. Número de accidentes en cada día de la semana

La segunda Xquery que hemos hecho consiste en calcular el número de accidentes que han ocurrido por cada día de la semana. Mostrándolo finalmente como un XML mostrando el día de la semana y el número de accidentes que se han producido. Esta query la hemos hecho con

el siguiente código:

```
xquery version "3.1";

let $days := distinct-values(
  for $x in doc("xml_bicycle_accidents.xml")/accidents/Accident
  return data($x/Week_day)
)

for $day in $days
return <results>
  <day>{$day}</day>
  <n_accidents>{
    count(
      for $x in doc("xml_bicycle_accidents.xml")/accidents/
      Accident/Week_day
      where $x = $day
      return $x
    )
  }</n_accidents>
</results>
```

Así la salida que hemos obtenido de la query es la siguiente:

```
<results>
  <day>Wednesday</day>
  <n_accidents>19</n_accidents>
</results>

<results>
  <day>Thursday</day>
  <n_accidents>16</n_accidents>
</results>

<results>
  <day>Tuesday</day>
  <n_accidents>17</n_accidents>
```

```

</results>

<results>
  <day>Sunday</day>
  <n_accidents>11</n_accidents>
</results>

<results>
  <day>Friday</day>
  <n_accidents>13</n_accidents>
</results>

<results>
  <day>Monday</day>
  <n_accidents>19</n_accidents>
</results>

<results>
  <day>Saturday</day>
  <n_accidents>12</n_accidents>
</results>

```

Así los días con mayor número de accidentes en nuestro dataset son los lunes y los miércoles.

### 2.2.3. Tipo de accidentes según cada condición de la carretera

En la tercera Xquery que hemos hecho, hemos pedido obtener las diferentes condiciones de la carretera en los accidentes producidos con en este código:

```

xquery version "3.1";
distinct-values(
for $cond in doc('/db/apps/doc/xml_bicycle_accidents.xml')/accidents/
  Accident/Conditions/Road_conditions
return data($cond)
)

```

Hemos obtenido este resultado:

"Wet"  
"Dry"  
"Snow"  
"Frost"

Como podemos ver, los accidentes en bici ocurren en cualquier condición de la carretera, por lo que podemos llegar a la conclusión de que suelen estar más influenciados por el comportamiento del ciclista o el conductor.

## **2.3. Transformaciones XSLT**

También hemos creado dos transformaciones XSLT, una que transforma el XML en un nuevo XML con los campos traducidos al español y una segunda transformación que transforma el XML a un HTML que muestra los datos en una tabla y es incrustable en una página web.

### **2.3.1. De XML a XML**

Con esta transformación buscamos traducir los atributos y etiquetas, que originalmente, estaban en inglés. También hemos hecho unos pequeños cambios en la estructura, la fecha ahora está unida en una sola etiqueta cuando antes estaba dividida en tres. A su vez hemos creado la etiqueta 'Detalles' para agrupar 'Fecha', 'Vehículos involucrados' y 'Damnificados' dentro de ella.



```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <Accidentes>
3    <Accidente ID="19954100C0101">
4      <Detalles>
5        <Fecha>Monday - 1995-01-16 - 12:30</Fecha>
6        <Vehículos_involucrados>2</Vehículos_involucrados>
7        <Damnificados>1</Damnificados>
8      </Detalles>
9      <Condiciones>
10       <Tipo_Via>Roundabout</Tipo_Via>
11       <Limite_Velocidad>30.0</Limite_Velocidad>
12       <Via_Estado/>
13       <Clima>Clear</Clima>
14       <Luminosidad>Daylight</Luminosidad>
15     </Condiciones>
16     <Ciclista>
17       <Género>Male</Género>
18       <Edad>26 to 35</Edad>
19       <Grado_Lesion>Slight</Grado_Lesion>
20     </Ciclista>
21   </Accidente>
22   .
23   .
24   .
25 </Accidentes>
```

Figura 2: XML transformation output

### 2.3.2. De XML a HTML

Accidentes de bicicletas en Reino Unido														
#	Index	N_vehicles	N_casualties	Date	Time	Week_day	Speed limit	Road Conditions	Road Type	Weather Conditions	Light Conditions	Biker gender	Biker age group	Biker severity Injury
1	19954100C0101	2	1	1995-01-16	12:30	Monday	30.0	Dry	Roundabout	Clear	Daylight	Male	26 to 35	Slight
2	19954100C0035	2	2	1995-01-19	08:15	Thursday	30.0	Frost	Single carriageway	Clear	Daylight	Female	56 to 65	Serious
3	19954100C0048	2	2	1995-01-21	11:38	Saturday	30.0	Wet	Roundabout	Rain and windy	Daylight	Male	16 to 20	Slight
4	19954100C0057	2	1	1995-01-22	09:20	Sunday	30.0	Wet	Single carriageway	Unknown	Daylight	Male	16 to 20	Slight
5	19954100B0043	2	1	1995-01-24	06:40	Tuesday	50.0	Dry	Single carriageway	Clear	Daylight	Male	26 to 35	Slight
6	19954100B0030	2	1	1995-01-25	22:45	Wednesday	30.0	Wet	Single carriageway	Clear and windy	Darkness lights lit	Male	26 to 35	Slight
7	19954100C0031	2	1	1995-01-26	17:30	Thursday	60.0	Wet	Single carriageway	Clear and windy	Darkness lights lit	Male	26 to 35	Slight
8	19954100C0067	2	2	1995-01-28	20:25	Saturday	30.0	Wet	Single carriageway	Rain	Darkness lights lit	Male	26 to 35	Slight
9	19954100B0058	2	1	1995-02-03	12:30	Thursday	30.0	Dry	Roundabout	Clear	Daylight	Female	21 to 25	Slight
10	19954100C0281	2	1	1995-02-05	08:30	Tuesday	60.0	Dry	Single carriageway	Clear	Daylight	Male	16 to 20	Slight

Figura 3: HTML transformation result



# 3

# Ontología

## 3.1. Grafo RDF

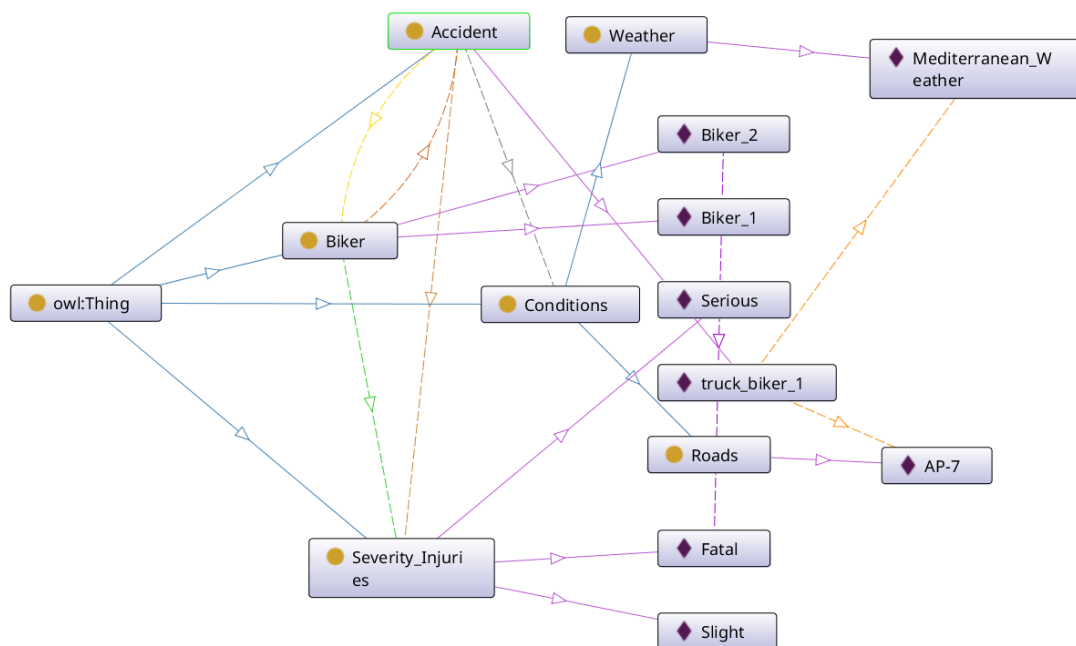


Figura 4: Ontograph



UNIVERSIDAD  
DE MÁLAGA

| **uma.es**

E.T.S. DE INGENIERÍA INFORMÁTICA

E.T.S de Ingeniería Informática  
Bulevar Louis Pasteur, 35  
Campus de Teatinos  
29071 Málaga