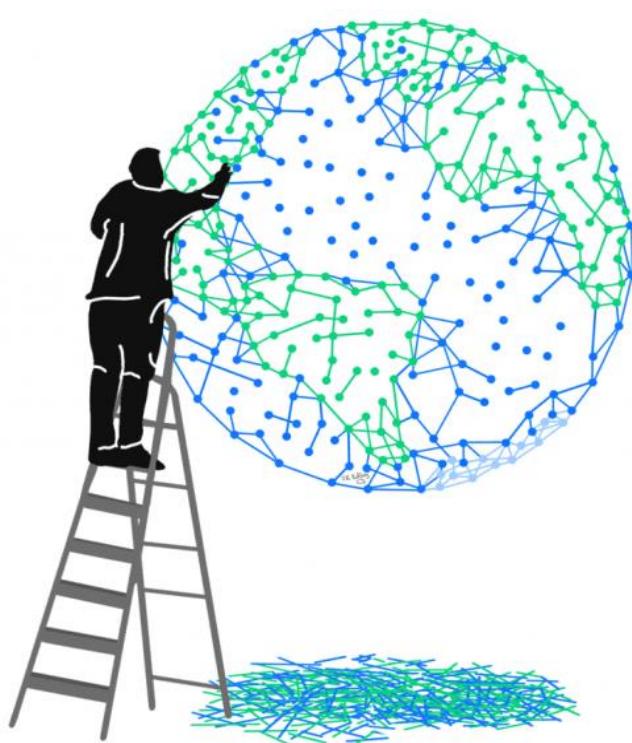
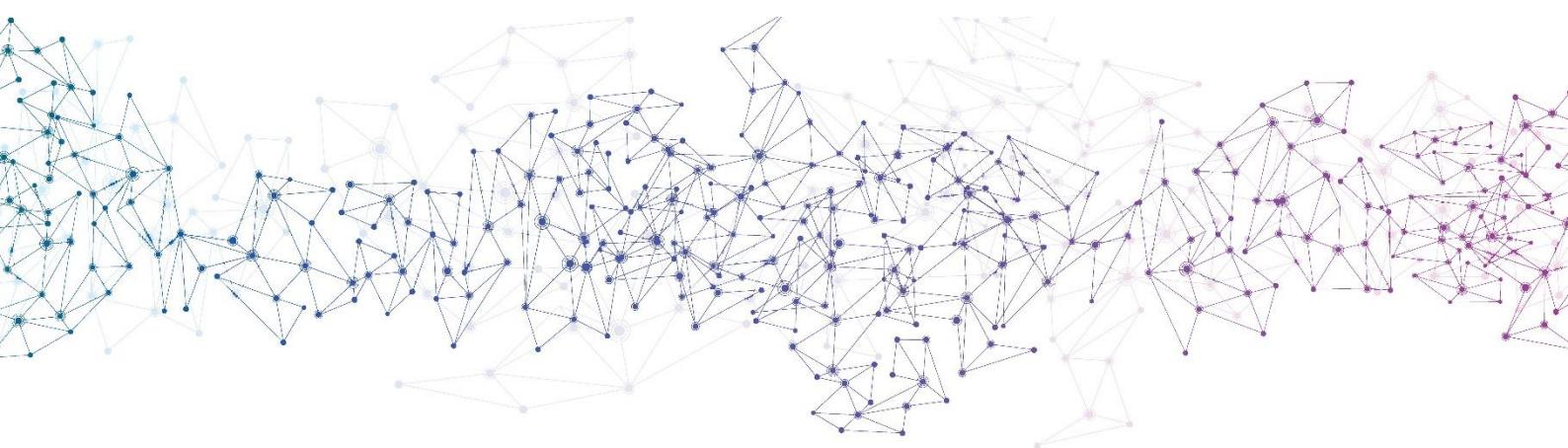




INFORME PRÁCTICA ENTREGABLE

Explotación Semántica de Datos

Máster en Bioinformática



ÍNDICE

1. Objetivo y función del RDF dataset.....	2
2. Desarrollo del RDF dataset.....	3
3. Dataset en el repositorio Blazegraph.....	8
4. Consultas en SPARQL desde Blazegraph	9
5. Consultas SPARQL desde RStudio	17
6. Publicación de los datos y aplicación de principios FAIR	21
7. Directorio de archivos generados	27

1. Objetivo y función del RDF dataset

La esclerosis múltiple (EM) es una enfermedad autoinmune del sistema nervioso central que afecta a la mielina, que es el recubrimiento protector de las fibras nerviosas. La etiología exacta de la enfermedad no se comprende completamente y se cree que es el resultado de una interacción compleja entre factores genéticos y ambientales.

Además, es muy heterogénea, es decir, afecta de diferentes maneras a diferentes individuos, lo que dificulta el desarrollo de enfoques de tratamiento y diagnóstico universalmente efectivos. De hecho, se reconocen diferentes formas de E.M. y cada una de ellas con sus propias características y progresión.

A lo anteriormente comentado, se le suma la falta de biomarcadores específicos para diagnosticar la enfermedad de manera definitiva y como consecuencia se puede retrasar el diagnóstico y el inicio del tratamiento.

Por tanto, la investigación sobre la esclerosis múltiple puede ser compleja debido a la naturaleza multifacética de la enfermedad y a la falta de comprensión completa de sus causas y mecanismos.

El objetivo del dataset realizado es unir conceptos sobre la enfermedad esclerosis múltiple con el fin de facilitar la recopilación de información cuando se lleve a cabo un estudio sobre la misma por investigadores.

Entre las vías de investigación que se encuentran actualmente abiertas encontramos: búsqueda de inmunomoduladores que controlen la respuesta autoinmune (inmunoterapia), búsqueda de enfoques para promover la reparación de la mielina y las células dañadas, exploración de terapias neuro-protectoras, búsqueda de biomarcadores e investigación de factores genéticos de riesgo.

En concreto, nos centraremos en la representación de importantes proteínas, genes, variantes y fármacos que están asociados de diferentes formas con la enfermedad y sus subtipos.

Algunos ejemplos son:

- **Proteínas:** interfieren en la respuesta autoinmune relacionada con los linfocitos T (“HLA-DRB1” y “CD86”), proteína esencial de la mielina (MBP) o son dianas de fármacos (“IFNAR1”, “S1PR1” y “ITGA4”).
- **Genes:** asociados a variantes de pacientes enfermos (“IL7R” o “TNFRSF1A”, entre otros).
- **Fármacos:** asociados a proteínas y a subtipos de enfermedad (“Fingolimod”, “Mitoxantrone” o “Natalizumab”).
- **Variantes:** variantes en pacientes con síntomas más severos (“rs7191888-G”) o variantes asociadas a la enfermedad EM independientemente de su severidad (“rs71314176-G”).
- **Enfermedad y subtipos:** “primary progressive multiple sclerosis”, “secondary progressive multiple sclerosis”, “progressive relapsing multiple sclerosis”, “relapsing-remitting multiple sclerosis”.

La idea de este modelo sería ampliarlo, ya que esto es solo una pequeña representación y hacerlo más preciso y detallado a medida que se vayan haciendo descubrimientos sobre la enfermedad, construyendo una gran red semántica que facilite el acceso a información tanto a investigadores como médicos dedicados al estudio de la esclerosis múltiple.

2. Desarrollo del RDF dataset

El desarrollo del dataset se ha llevado a cabo con uno de los modelos estándar de intercambio de datos en la web, en concreto, el modelo usado es “Resource Description Framework” (RDF).

El modelo de datos RDF se basa en relacionar entidades o recursos, por medio de relaciones binarias o enunciados. Cada enunciado es una tripleta compuesta por tres elementos, **sujeto-predicado-objeto**. Estos tres elementos en conjunto se denominan **tripleta** y hace referencia a:

Sujeto: se trata del recurso o entidad descrito.

Predicado o propiedad: relación o propiedad del recurso que se está describiendo.

Objeto: valor concreto de la propiedad que describe al recurso (puede ser un nuevo recurso o un literal).

RDF es el lenguaje recomendado por el “World Wide Web consortium” (W3C) para representar los recursos y objetos en la Web Semántica. Proporciona interoperabilidad entre aplicaciones que intercambian información legible por la máquina en la Web. RDF destaca por la facilidad para habilitar el procesamiento automatizado de los recursos Web. Es capaz de ampliar la red aumentando su estructura por medio de URIs, donde cada parte de cada tripleta tendrá su URI correspondiente.

Un conjunto de tripletas se denomina grafo y es precisamente lo que hemos ido construyendo tripleta a tripleta. Para ello hemos usado diferentes fuentes de información para incluir distintos recursos como **Uniprot**, **NCBI**, **DrugBank**, entre otros. Además, hemos hecho uso de distintos repositorios de ontologías como **GeneOntology** y **OboFoundry** que nos proporcionan un vocabulario adicional que proporcionan al grafo un mayor enriquecimiento semántico.

El modelo RDF presenta varias opciones de sintaxis, la usada para el desarrollo de este dataset es “Turtle Terse RDF Triple Language” y más adelante será convertido a sintaxis RDF/XML por medio de la herramienta <http://www.easyrdf.org/converter>

En primera instancia, procedemos a definir los namespaces empleados (será posible abreviar o expandir mediante el uso de prefijos y estableceremos los siguientes:

- **Vocabularios reservados** (definen lenguajes)

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.  
@prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
```

- **RDF de UniprotKB**

```
@prefix uniprot:<http://purl.uniprot.org/uniprot/>.  
@prefix uniprott:<http://purl.uniprot.org/core/>.  
@prefix up:<https://www.uniprot.org/core/>.
```

- **Recursos de ontologías**

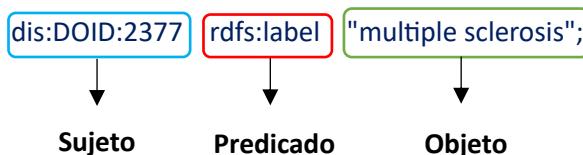
```
@prefix go: <http://amigo.geneontology.org/amigo/term/>.  
@prefix dis:<https://www.disease-ontology.org/do/>.  
@prefix onto:<http://purl.obolibrary.org/obo/>.
```

- **Otros recursos**

```
@prefix gen:<https://www.ncbi.nlm.nih.gov/gene/>.
@prefix far:<https://go.drugbank.com/drugs/>.
@prefix var:<https://www.ebi.ac.uk/gwas/variants/>.
@prefix pub:<https://pubmed.ncbi.nlm.nih.gov/>.
```

Una vez definidos los prefijos, el siguiente paso es ir escribiendo el grafo de triplete en triplete siguiendo la idea de relaciones anteriormente comentada.

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix dis:<https://www.disease-ontology.org/do/>.
```



Para continuar asociando predicados y objetos a un mismo sujeto se lleva a cabo la serialización de los sujetos, añadiendo todas las propiedades y objetos deseados antes de pasar al siguiente sujeto.

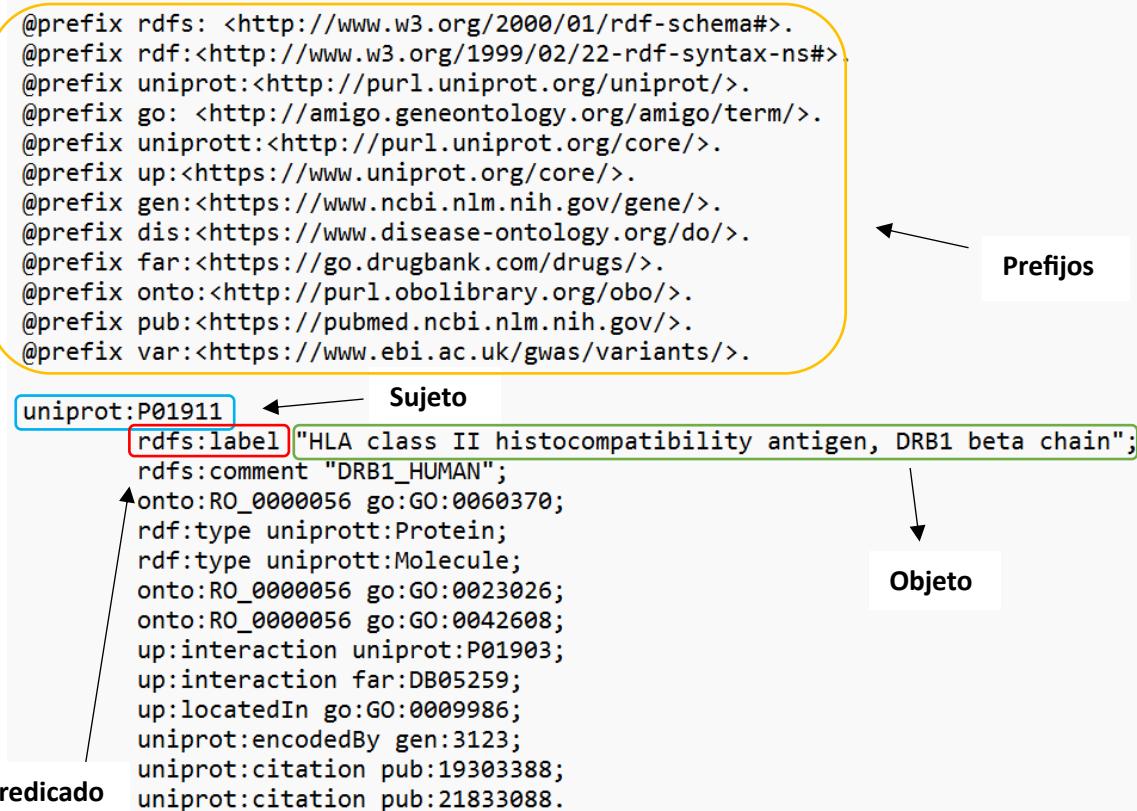


Figura 1.- Tripletas grafo RDF turtle (archivo: tripletas_entrega.txt)

Al convertir nuestro grafo a sintaxis RDF/XML nos salen un total de 349 tripletas.

El próximo será usar la herramienta descrita para convertir nuestro grafo en sintaxis turtle a sintaxis RDF/XML y a continuación usaremos la herramienta <https://www.w3.org/RDF/Validator/> para validar nuestro grafo RDF y ver si es correcto.

De la siguiente manera se puede visualizar una parte del grafo y sus tripletas correspondiente a la proteína Interferon alpha/beta receptor 1 con ID de uniprot: P17181.

Triples of the Data Model			
Number	Subject	Predicate	Object
1	http://purl.uniprot.org/uniprot/P17181	http://www.w3.org/2000/01/rdf-schema#label	"Interferon alpha/beta receptor 1"
2	http://purl.uniprot.org/uniprot/P17181	http://www.w3.org/2000/01/rdf-schema#comment	"INAR1_HUMAN"
3	http://purl.uniprot.org/uniprot/P17181	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://purl.uniprot.org/core/Protein
4	http://purl.uniprot.org/uniprot/P17181	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://purl.uniprot.org/core/Molecule
5	http://purl.uniprot.org/uniprot/P17181	http://purl.obolibrary.org/obo/RO_0000056	http://amigo.geneontology.org/amigo/term/GO:0035457
6	http://purl.uniprot.org/uniprot/P17181	http://purl.obolibrary.org/obo/RO_0000056	http://amigo.geneontology.org/amigo/term/GO:0035458
7	http://purl.uniprot.org/uniprot/P17181	http://purl.obolibrary.org/obo/RO_0000056	http://amigo.geneontology.org/amigo/term/GO:0007259
8	http://purl.uniprot.org/uniprot/P17181	http://purl.obolibrary.org/obo/RO_0000056	http://amigo.geneontology.org/amigo/term/GO:0004905
9	http://purl.uniprot.org/uniprot/P17181	http://purl.obolibrary.org/obo/RO_0000056	http://amigo.geneontology.org/amigo/term/GO:0008269
10	http://purl.uniprot.org/uniprot/P17181	https://www.uniprot.org/core/interaction	http://purl.uniprot.org/uniprot/P2597
11	http://purl.uniprot.org/uniprot/P17181	https://www.uniprot.org/core/interaction	https://go.drugbank.com/drugs/DB00068
12	http://purl.uniprot.org/uniprot/P17181	https://www.uniprot.org/core/interaction	https://go.drugbank.com/drugs/DB09122
13	http://purl.uniprot.org/uniprot/P17181	https://www.uniprot.org/core/locatedIn	https://go.drugbank.com/drugs/DB14999
14	http://purl.uniprot.org/uniprot/P17181	https://www.uniprot.org/core/locatedIn	http://amigo.geneontology.org/amigo/term/GO:0005886
15	http://purl.uniprot.org/uniprot/P17181	http://purl.uniprot.org/encodedBy	https://www.ncbi.nlm.nih.gov/gene/3454
16	http://purl.uniprot.org/uniprot/P17181	http://purl.uniprot.org/uniprot/citation	https://pubmed.ncbi.nlm.nih.gov/2153461
17	http://purl.uniprot.org/uniprot/P17181	http://purl.uniprot.org/uniprot/citation	https://pubmed.ncbi.nlm.nih.gov/7813427

Figura 2.- Tripletas RDF/XML P17181 validado

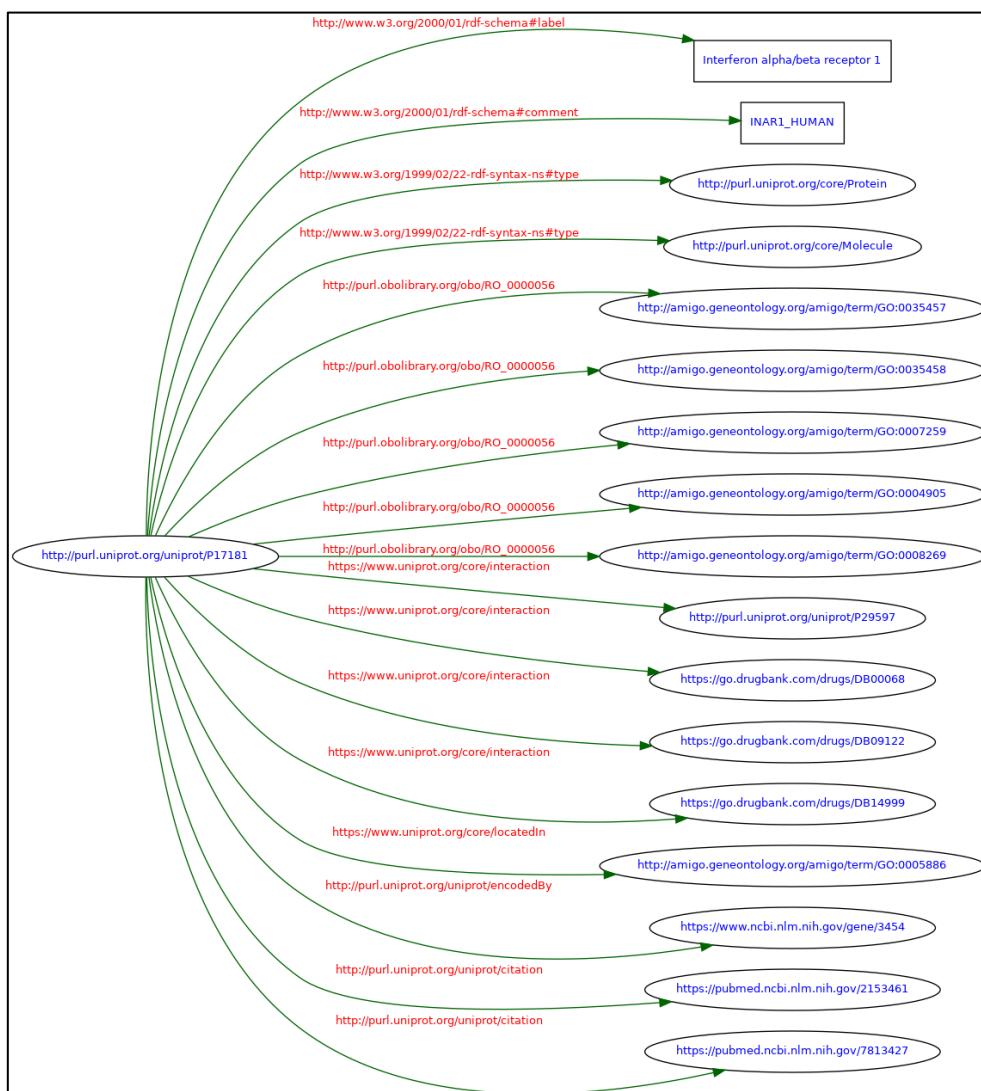


Figura 3.- Grafo RDF/XML P17181

Podemos ver en el grafo que este sujeto forma varias tripletas, teniendo múltiples objetos como serían: una etiqueta, un comentario, varios procesos biológicos, varias funciones moleculares, una proteína con la que interacciona, tres fármacos con los que interacciona, el gen que codifica la proteína y dos publicaciones en las que aparece (figura 3).

A partir de esta proteína y de sus objetos, que a su vez serán sujetos de otras tripletas (igual que esta proteína será objeto de otras tripletas), se van a ir relacionando con predicados y con otros objetos y se irán llevando a cabo todas las tripletas del grafo y, por tanto, del dataset.

Las relaciones y propiedades se han adquirido de las siguientes fuentes de información:

- **Proteínas:** <https://www.uniprot.org/> donde se han buscado asociaciones de proteínas tanto con otras proteínas, con fármacos, el gen que las codifica,
- **GO:** <http://amigo.geneontology.org/amigo/term/> procesos biológicos, funciones moleculares y componente celular donde se sitúan proteínas.
- **Genes:** <https://www.ncbi.nlm.nih.gov/gene/> búsqueda de genes que codifican proteínas de interés y que poseen variantes de riesgo de la enfermedad.
- **Fármacos:** <https://go.drugbank.com/drugs/> donde se ha buscado asociaciones de fármacos con la enfermedad y con proteínas de interés.
- **Enfermedades:** <https://www.disease-ontology.org/do/> búsqueda de categorías o clases y subclases en las que se encuentra la EM y si ella misma tiene subtipos definidos.
- **Variantes:** <https://www.ebi.ac.uk/gwas/variants/> donde se han buscado asociaciones de genes con variantes de riesgo asociadas la enfermedad EM y variantes asociadas con severidad de la misma.
- **Ontologías:** <http://purl.obolibrary.org/obo/> ha servido como soporte para buscar toda clase de ontologías junto con <http://purl.uniprot.org/core/>
- **Publicaciones:** <https://pubmed.ncbi.nlm.nih.gov/> únicamente correspondientes a proteínas.

A continuación, al cambiar nuestro grafo de sintaxis vemos que está formado por 349 tripletas y en la figura 5 podemos ver su representación donde se relación la mayoría de tripletas entre sí cumpliendo con el objetivo de la generación de ontologías y de grafos de conocimiento en la Web Semántica.

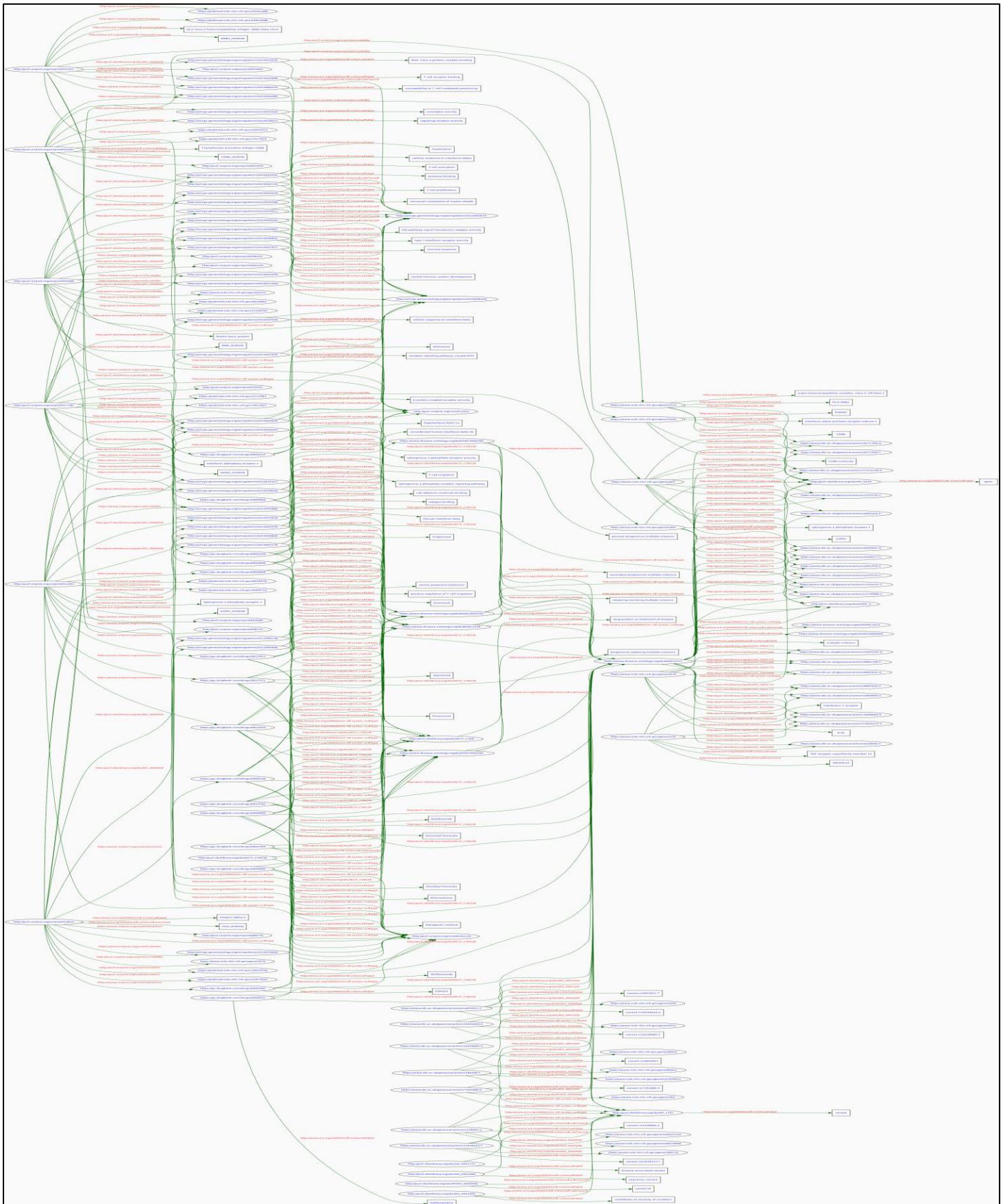
Output

Number of triples parsed: 349

```
<?xml version="1.0" encoding="utf-8" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:ns0="http://purl.obolibrary.org/obo/"
    xmlns:ns1="https://www.uniprot.org/core/"
    xmlns:ns2="http://purl.uniprot.org/uniprot/"

    <rdf:Description rdf:about="http://purl.uniprot.org/uniprot/P01911">
        <rdfs:label>HLA class II histocompatibility antigen, DRB1 beta chain</rdfs:label>
        <rdfs:comment>DRB1_HUMAN</rdfs:comment>
        <ns0:RO_0000056>
            <rdf:Description rdf:about="http://amigo.geneontology.org/amigo/term/GO:0060370">
                <rdfs:subClassOf rdf:resource="http://amigo.geneontology.org/amigo/term/GO:0008150"/>
                <rdfs:label>susceptibility to T cell mediated cytotoxicity</rdfs:label>
            </rdf:Description>
        </ns0:RO_0000056>
```

Figura 4.- Conversión Turtle a RDF/XML

**Figura 5.- Grafo RDF completo formado por 349 triplets**

3. Dataset en el repositorio Blazegraph

Una vez construido el código de nuestro dataset y convertido desde Turtle a RDF/XML, procedemos a subirlo al repositorio Blazegraph, que en nuestro caso será ejecutado desde GoogleCloud a través de nuestra instancia VM mediante el contenido del repositorio github de Mikel Egana en SSH. El objetivo es poder realizar consultas a nuestro código en lenguaje SPARQL.

```
franciscojavierlopezcarbonell@miservidor:~$ cd UM-Bioinformatics-MSc-FAIR-data/LinkedDataServer/blazegraph/
franciscojavierlopezcarbonell@miservidor:~/UM-Bioinformatics-MSc-FAIR-data/LinkedDataServer/blazegraph$ java -se
rver -Djetty.port=3031 -jar blazegraph.jar
```

```
Welcome to the Blazegraph(tm) Database.

Go to http://10.128.0.2:3031/blazegraph/ to get started.
```

A continuación, copiamos nuestro código en formato RDF/XML (figura 6 “update”), especificando tipo de código y formato antes de actualizar (figura 7).

Actualizado nuestro código RDF ya podemos comenzar a realizar nuestras consultas SPARQL (figura 6 “query”).

```
477 <rdf:Description rdf:about="https://go.drugbank.com/drugs/DB06637">
478   <rdfs:label>Dalfampridine</rdfs:label>
479   <rdf:type rdf:resource="http://purl.uniprot.org/core/Molecule"/>
480   <rdf:type rdf:resource="http://purl.obolibrary.org/obo/NCIT_C459"/>
481   <ns0:NCIT_C49236 rdf:resource="https://www.disease-ontology.org/do/DOID:2377"/>
482 </rdf:Description>
```

Figura 6.- Dataset RDF/XML cargado en Blazegraph

Figura 7.- Actualización dataset RDF/XML en Blazegraph

4. Consultas en SPARQL desde Blazegraph

El lenguaje de consulta SPARQL es una especificación propuesta por el W3C, que proporciona una sintaxis para manipular grafos RDF en la web y en repositorios de datos. La estructura básica de una consulta SPARQL está compuesta de los siguientes elementos:

- **Prefijos:** evitan el uso de URIs completas para acceder a los datos
- **Select:** usado para seleccionar las variables que queremos recibir
- **Where:** delimita el patron del grafo que queremos extraer de nuestro grafo mayor.
- **Otros:** podemos modificar nuestras consultas usando “**group by**”, “**order by**”, “**limit**” para agrupar, ordenar o limitar nuestro conjunto de datos a consultar. Además, podemos realizar otro tipo de consultas preguntando por un valor verdadero o falso si es correcta o falsa la estructura RDF consultada (“**Ask**”) o solicitando una descripción de una estructura RDF indicada (“**Describe**”).

Consulta 1: búsqueda de todas las tripletas (sujeto, predicado y objeto) que forman el dataset.

	Time	Query	Results	Execution Time	Delete
1 SELECT * 2 WHERE { 3 ?s ?p ?o 4 }	2023-12-09T13:14:38.415Z	SELECT * WHERE { ?s ?p ?o }	349	1sec, 238ms	X

s	p	o
<https://www.ebi.ac.uk/gwas/variants/rs180358-T>	<http://purl.obolibrary.org/obo/PORO_0000946>	<https://www.ncbi.nlm.nih.gov/gene/84811>
<https://www.ebi.ac.uk/gwas/variants/rs180358-T>	<http://purl.obolibrary.org/obo/R0_0003305>	<https://www.disease-ontology.org/do/DOID:2377>
<https://www.ebi.ac.uk/gwas/variants/rs180358-T>	rdf:type	<http://purl.obolibrary.org/obo/MI_1241>
<https://www.ebi.ac.uk/gwas/variants/rs180358-T>	rdfs:label	variant rs180358-T
<https://www.ebi.ac.uk/gwas/variants/rs4953911-T>	<http://purl.obolibrary.org/obo/PORO_0000946>	<https://www.ncbi.nlm.nih.gov/gene/4249>
<https://www.ebi.ac.uk/gwas/variants/rs4953911-T>	<http://purl.obolibrary.org/obo/R0_0003305>	<https://www.disease-ontology.org/do/DOID:2377>
<https://www.ebi.ac.uk/gwas/variants/rs4953911-T>	rdf:type	<http://purl.obolibrary.org/obo/MI_1241>
<https://www.ebi.ac.uk/gwas/variants/rs4953911-T>	rdfs:label	variant rs4953911-T
<https://www.ebi.ac.uk/gwas/variants/rs7191888-G>	<http://purl.obolibrary.org/obo/PORO_0000946>	<https://www.ncbi.nlm.nih.gov/gene/463>
<https://www.ebi.ac.uk/gwas/variants/rs7191888-G>	<http://purl.obolibrary.org/obo/R0_0003305>	<https://www.disease-ontology.org/do/DOID:2377>
<https://www.ebi.ac.uk/gwas/variants/rs7191888-G>	rdf:type	<http://purl.obolibrary.org/obo/MI_1241>
<https://www.ebi.ac.uk/gwas/variants/rs7191888-G>	rdfs:label	variant rs7191888-G
<https://www.ncbi.nlm.nih.gov/gene/1901>	<http://purl.obolibrary.org/obo/SO_0001060>	<https://www.ebi.ac.uk/gwas/variants/rs12756986-C>
<https://www.ncbi.nlm.nih.gov/gene/1901>	<http://purl.obolibrary.org/obo/SO_0001060>	<https://www.ebi.ac.uk/gwas/variants/rs17123757-C>
<https://www.ncbi.nlm.nih.gov/gene/1901>	rdf:type	<http://purl.obolibrary.org/obo/MI_0250>
<https://www.ncbi.nlm.nih.gov/gene/1901>	rdfs:comment	S1PR1
<https://www.ncbi.nlm.nih.gov/gene/1901>	rdfs:label	sphingosine-1-phosphate receptor 1
<https://www.ncbi.nlm.nih.gov/gene/3123>	<http://purl.obolibrary.org/obo/SO_0001060>	<https://www.ebi.ac.uk/gwas/variants/rs9271366-G>
<https://www.ncbi.nlm.nih.gov/gene/3123>	<http://purl.obolibrary.org/obo/SO_0001060>	<https://www.ebi.ac.uk/gwas/variants/rs9271640-T>
<https://www.ncbi.nlm.nih.gov/gene/3123>	rdf:type	<http://purl.obolibrary.org/obo/MI_0250>
<https://www.ncbi.nlm.nih.gov/gene/3123>	rdfs:comment	HLA-DRB1
<https://www.ncbi.nlm.nih.gov/gene/3123>	rdfs:label	major histocompatibility complex, class II, DR beta 1

Podemos apreciar 349 resultados de consulta, que son exactamente el número de tripletas que conforman nuestro RDF dataset y vemos seguidamente un extracto de algunos de los sujetos, predicados y objetos que las forman.

Consulta 2: devuelve todas las variantes génicas en secuencia asociadas directamente a genes, su etiqueta y los genes que la contienen.

```

1 prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
2 prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 prefix uniprot:<http://purl.uniprot.org/uniprot/>
4 prefix go: <http://amigo.geneontology.org/amigo/term/>
5 prefix uniprott:<http://purl.uniprot.org/core/>
6 prefix up:<https://www.uniprot.org/core/>
7 prefix gen:<https://www.ncbi.nlm.nih.gov/gene/>
8 prefix dis:<https://www.disease-ontology.org/do/>
9 prefix far:<https://go.drugbank.com/drugs/>
10 prefix onto:<http://purl.obolibrary.org/obo/>
11 prefix pub:<https://pubmed.ncbi.nlm.nih.gov/>
12 prefix var:<https://www.ebi.ac.uk/gwas/variants/>
13
14 SELECT ?var ?gen ?label
15 WHERE {
16   ?var onto:PORO_0000946 ?gen.
17   ?var rdfs:label ?label
18 }
19

```

var	gen	label
< https://www.ebi.ac.uk/gwas/variants/rs10243024-A >	< https://www.ncbi.nlm.nih.gov/gene/4233 >	variant rs10243024-A
< https://www.ebi.ac.uk/gwas/variants/rs10259085-C >	< https://www.ncbi.nlm.nih.gov/gene/56913 >	variant rs10259085-C
< https://www.ebi.ac.uk/gwas/variants/rs12638253-T >	< https://www.ncbi.nlm.nih.gov/gene/389170 >	variant rs12638253-T
< https://www.ebi.ac.uk/gwas/variants/rs1478091-C >	< https://www.ncbi.nlm.nih.gov/gene/100421232 >	variant rs1478091-C
< https://www.ebi.ac.uk/gwas/variants/rs1478091-C >	< https://www.ncbi.nlm.nih.gov/gene/106479666 >	variant rs1478091-C
< https://www.ebi.ac.uk/gwas/variants/rs180358-T >	< https://www.ncbi.nlm.nih.gov/gene/101929011 >	variant rs180358-T
< https://www.ebi.ac.uk/gwas/variants/rs180358-T >	< https://www.ncbi.nlm.nih.gov/gene/84811 >	variant rs180358-T
< https://www.ebi.ac.uk/gwas/variants/rs4953911-T >	< https://www.ncbi.nlm.nih.gov/gene/4249 >	variant rs4953911-T
< https://www.ebi.ac.uk/gwas/variants/rs7191888-G >	< https://www.ncbi.nlm.nih.gov/gene/463 >	variant rs7191888-G

Del mismo modo, procedemos a buscar proteínas (junto con su comentario) que interaccionen con fármacos (junto su etiqueta).

```

1 prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
2 prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 prefix uniprot:<http://purl.uniprot.org/uniprot/>
4 prefix go: <http://amigo.geneontology.org/amigo/term/>
5 prefix uniprott:<http://purl.uniprot.org/core/>
6 prefix up:<https://www.uniprot.org/core/>
7 prefix gen:<https://www.ncbi.nlm.nih.gov/gene/>
8 prefix dis:<https://www.disease-ontology.org/do/>
9 prefix far:<https://go.drugbank.com/drugs/>
10 prefix onto:<http://purl.obolibrary.org/obo/>
11 prefix pub:<https://pubmed.ncbi.nlm.nih.gov/>
12 prefix var:<https://www.ebi.ac.uk/gwas/variants/>
13
14 SELECT ?protein ?far ?comment ?label
15 WHERE {
16   ?protein up:interaction ?far.
17   ?protein rdfs:comment ?comment.
18   ?far rdfs:label ?label
19 }
20

```

protein	far	comment	label
http://purl.uniprot.org/uniprot/P17181	https://go.drugbank.com/drugs/DB00068	INAR1_HUMAN	recombinant human Interferon beta-1b
http://purl.uniprot.org/uniprot/P13612	https://go.drugbank.com/drugs/DB00108	ITA4_HUMAN	Natalizumab
http://purl.uniprot.org/uniprot/P13612	https://go.drugbank.com/drugs/DB05092	ITA4_HUMAN	CDP323
http://purl.uniprot.org/uniprot/P01911	https://go.drugbank.com/drugs/DB05259	DRB1_HUMAN	Glatiramer
http://purl.uniprot.org/uniprot/P21453	https://go.drugbank.com/drugs/DB08888	S1PR1_HUMAN	Fingolimod
http://purl.uniprot.org/uniprot/P17181	https://go.drugbank.com/drugs/DB09122	INAR1_HUMAN	Peginterferon beta-1a
http://purl.uniprot.org/uniprot/P21453	https://go.drugbank.com/drugs/DB12016	S1PR1_HUMAN	Ponesimod
http://purl.uniprot.org/uniprot/P21453	https://go.drugbank.com/drugs/DB12371	S1PR1_HUMAN	Siponimod
http://purl.uniprot.org/uniprot/P21453	https://go.drugbank.com/drugs/DB12612	S1PR1_HUMAN	Ozanimod
http://purl.uniprot.org/uniprot/P17181	https://go.drugbank.com/drugs/DB14999	INAR1_HUMAN	Human interferon beta

Consulta 3: buscamos proteínas que se haya registrado el gen que las codifica, genes que tienen variantes de secuencia asociadas (pedimos esas variantes también) y enfermedades que tienen variantes asociadas (pedimos esas variantes también).

```

1 prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
2 prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 prefix uniprot:<http://purl.uniprot.org/uniprot/>
4 prefix go: <http://amigo.geneontology.org/amigo/term/>
5 prefix uniprott:<http://purl.uniprot.org/core/>
6 prefix up:<https://www.uniprot.org/core/>
7 prefix gen:<https://www.ncbi.nlm.nih.gov/gene/>
8 prefix dis:<https://www.disease-ontology.org/do/>
9 prefix far:<https://go.drugbank.com/drugs/>
10 prefix onto:<http://purl.obolibrary.org/obo/>
11 prefix pub:<https://pubmed.ncbi.nlm.nih.gov/>
12 prefix var:<https://www.ebi.ac.uk/gwas/variants/>
13
14 SELECT ?protein ?gen ?var ?dis
15 WHERE {
16   ?protein uniprot:encodedBy ?gen.
17   ?gen onto:SO_0001060 ?var.
18   ?dis onto:SO_0001771 ?var
19 }
20

```

protein	gen	var	dis
http://purl.uniprot.org/uniprot/P21453	https://www.ncbi.nlm.nih.gov/gene/1901	https://www.ebi.ac.uk/gwas/variants/rs12756986-C	https://www.disease-ontology.org/do/DOID:2377
http://purl.uniprot.org/uniprot/P21453	https://www.ncbi.nlm.nih.gov/gene/1901	https://www.ebi.ac.uk/gwas/variants/rs17123757-C	https://www.disease-ontology.org/do/DOID:2377
http://purl.uniprot.org/uniprot/P42081	https://www.ncbi.nlm.nih.gov/gene/942	https://www.ebi.ac.uk/gwas/variants/rs2255214-C	https://www.disease-ontology.org/do/DOID:2377
http://purl.uniprot.org/uniprot/P42081	https://www.ncbi.nlm.nih.gov/gene/942	https://www.ebi.ac.uk/gwas/variants/rs2255214-T	https://www.disease-ontology.org/do/DOID:2377
http://purl.uniprot.org/uniprot/P42081	https://www.ncbi.nlm.nih.gov/gene/942	https://www.ebi.ac.uk/gwas/variants/rs2681424-C	https://www.disease-ontology.org/do/DOID:2377
http://purl.uniprot.org/uniprot/P42081	https://www.ncbi.nlm.nih.gov/gene/942	https://www.ebi.ac.uk/gwas/variants/rs2681424-T	https://www.disease-ontology.org/do/DOID:2377
http://purl.uniprot.org/uniprot/P42081	https://www.ncbi.nlm.nih.gov/gene/942	https://www.ebi.ac.uk/gwas/variants/rs4308217-C	https://www.disease-ontology.org/do/DOID:2377
http://purl.uniprot.org/uniprot/P17181	https://www.ncbi.nlm.nih.gov/gene/3454	https://www.ebi.ac.uk/gwas/variants/rs71314176-G	https://www.disease-ontology.org/do/DOID:2377
http://purl.uniprot.org/uniprot/P01911	https://www.ncbi.nlm.nih.gov/gene/3123	https://www.ebi.ac.uk/gwas/variants/rs9271366-G	https://www.disease-ontology.org/do/DOID:2377
http://purl.uniprot.org/uniprot/P01911	https://www.ncbi.nlm.nih.gov/gene/3123	https://www.ebi.ac.uk/gwas/variants/rs9271640-T	https://www.disease-ontology.org/do/DOID:2377
http://purl.uniprot.org/uniprot/P42081	https://www.ncbi.nlm.nih.gov/gene/942	https://www.ebi.ac.uk/gwas/variants/rs9282641-G	https://www.disease-ontology.org/do/DOID:2377

Consulta 4: vamos a pedir que nos cuente el número de fármacos, registrados como tipo tratamiento de enfermedad usando el operador “Count”. Nos aparece el siguiente resultado.

numero_far
15

```

1 prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
2 prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 prefix uniprot:<http://purl.uniprot.org/uniprot/>
4 prefix go: <http://amigo.geneontology.org/amigo/term/>
5 prefix uniprott:<http://purl.uniprot.org/core/>
6 prefix up:<https://www.uniprot.org/core/>
7 prefix gen:<https://www.ncbi.nlm.nih.gov/gene/>
8 prefix dis:<https://www.disease-ontology.org/do/>
9 prefix far:<https://go.drugbank.com/drugs/>
10 prefix onto:<http://purl.obolibrary.org/obo/>
11 prefix pub:<https://pubmed.ncbi.nlm.nih.gov/>
12 prefix var:<https://www.ebi.ac.uk/gwas/variants/>
13
14 SELECT (COUNT(?far) AS ?numero_far)
15 WHERE {
16   ?far rdf:type onto:NCIT_C459
17 }
18

```

A continuación, buscamos fármacos sean tratamiento de enfermedad (junto su etiqueta) y la enfermedad que tratan, usando el operador “**Filter**” para filtrar por tipo de enfermedad recurrente dentro de los subtipos de esclerosis múltiple registrados.

```

1 prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
2 prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 prefix uniprot:<http://purl.uniprot.org/uniprot/>
4 prefix go: <http://amigo.geneontology.org/amigo/term/>
5 prefix uniprott:<http://purl.uniprot.org/core/>
6 prefix up:<https://www.uniprot.org/core/>
7 prefix gen:<https://www.ncbi.nlm.nih.gov/gene/>
8 prefix dis:<https://www.disease-ontology.org/do/>
9 prefix far:<https://go.drugbank.com/drugs/>
10 prefix onto:<http://purl.obolibrary.org/obo/>
11 prefix pub:<https://pubmed.ncbi.nlm.nih.gov/>
12 prefix var:<https://www.ebi.ac.uk/gwas/variants/>
13
14 SELECT ?far ?label ?n
15 WHERE {
16   ?far onto:NCIT_C49236 ?dis.
17   ?far rdfs:label ?label.
18   ?dis rdfs:label ?n.
19   FILTER regex (?n, 'relapsing.*')
20 }

```

Podemos ver que existen dos tipos de E. múltiple recurrente y los fármacos que se asocian como tratamiento de las mismas aportando su URI y su etiqueta.

far	label	n
https://go.drugbank.com/drugs/DB00108	Natalizumab	progressive relapsing multiple sclerosis
https://go.drugbank.com/drugs/DB01204	Mitoxantrone	progressive relapsing multiple sclerosis
https://go.drugbank.com/drugs/DB08880	Teriflunomide	progressive relapsing multiple sclerosis
https://go.drugbank.com/drugs/DB08908	Dimethyl fumarate	progressive relapsing multiple sclerosis
https://go.drugbank.com/drugs/DB12016	Ponesimod	progressive relapsing multiple sclerosis
https://go.drugbank.com/drugs/DB12371	Siponimod	progressive relapsing multiple sclerosis
https://go.drugbank.com/drugs/DB12612	Ozanimod	progressive relapsing multiple sclerosis
https://go.drugbank.com/drugs/DB14783	Diroximel fumarate	progressive relapsing multiple sclerosis
https://go.drugbank.com/drugs/DB14999	Human interferon beta	progressive relapsing multiple sclerosis
https://go.drugbank.com/drugs/DB00068	recombinant human Interferon beta-1b	relapsing-remitting multiple sclerosis
https://go.drugbank.com/drugs/DB00108	Natalizumab	relapsing-remitting multiple sclerosis
https://go.drugbank.com/drugs/DB01204	Mitoxantrone	relapsing-remitting multiple sclerosis
https://go.drugbank.com/drugs/DB05259	Glatiramer	relapsing-remitting multiple sclerosis
https://go.drugbank.com/drugs/DB08868	Fingolimod	relapsing-remitting multiple sclerosis
https://go.drugbank.com/drugs/DB08880	Teriflunomide	relapsing-remitting multiple sclerosis
https://go.drugbank.com/drugs/DB08908	Dimethyl fumarate	relapsing-remitting multiple sclerosis
https://go.drugbank.com/drugs/DB12016	Ponesimod	relapsing-remitting multiple sclerosis
https://go.drugbank.com/drugs/DB12371	Siponimod	relapsing-remitting multiple sclerosis
https://go.drugbank.com/drugs/DB12612	Ozanimod	relapsing-remitting multiple sclerosis
https://go.drugbank.com/drugs/DB14783	Diroximel fumarate	relapsing-remitting multiple sclerosis
https://go.drugbank.com/drugs/DB14999	Human interferon beta	relapsing-remitting multiple sclerosis

Por último, en este apartado vamos a incluir una consulta sobre fármacos más específica, donde buscaremos aquellos que traten E. múltiple y sean tipo tratamiento y proteína.

```

1 prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
2 prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 prefix uniprot:<http://purl.uniprot.org/uniprot/>
4 prefix go: <http://amigo.geneontology.org/amigo/term/>
5 prefix uniprott:<http://purl.uniprot.org/core/>
6 prefix up:<https://www.uniprot.org/core/>
7 prefix gen:<https://www.ncbi.nlm.nih.gov/gene/>
8 prefix dis:<https://www.disease-ontology.org/do/>
9 prefix far:<https://go.drugbank.com/drugs/>
10 prefix onto:<http://purl.obolibrary.org/obo/>
11 prefix pub:<https://pubmed.ncbi.nlm.nih.gov/>
12 prefix var:<https://www.ebi.ac.uk/gwas/variants/>
13
14 SELECT ?far ?label
15 WHERE {
16   ?far rdf:type onto:NCIT_C459.
17   ?far rdfs:label ?label.
18   ?far onto:NCIT_C49236 dis:D0ID:2377.
19   ?far rdf:type uniprott:Protein
20
21 }
22

```

far	label
https://go.drugbank.com/drugs/DB00068	recombinant human Interferon beta-1b
https://go.drugbank.com/drugs/DB00108	Natalizumab
https://go.drugbank.com/drugs/DB09122	Peginterferon beta-1a
https://go.drugbank.com/drugs/DB14999	Human interferon beta

Consulta 5: pediremos cualquier entidad que tenga un comentario y/o una etiqueta (hacemos uso del operador “Union” para unir dos conjuntos de estructuras) y queremos que guarde relación con las células T, por lo que filtramos.

```

1 prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
2 prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 prefix uniprot:<http://purl.uniprot.org/uniprot/>
4 prefix go: <http://amigo.geneontology.org/amigo/term/>
5 prefix uniprott:<http://purl.uniprot.org/core/>
6 prefix up:<https://www.uniprot.org/core/>
7 prefix gen:<https://www.ncbi.nlm.nih.gov/gene/>
8 prefix dis:<https://www.disease-ontology.org/do/>
9 prefix far:<https://go.drugbank.com/drugs/>
10 prefix onto:<http://purl.obolibrary.org/obo/>
11 prefix pub:<https://pubmed.ncbi.nlm.nih.gov/>
12 prefix var:<https://www.ebi.ac.uk/gwas/variants/>
13
14 SELECT ?p ?n
15 WHERE {
16   {?p rdfs:label ?n}
17   UNION
18   {?p rdfs:comment ?n}
19   FILTER regex (?n, 'T cell.*')
20 }
21

```

P	n
http://amigo.geneontology.org/amigo/term/GO:2000406	positive regulation of T cell migration
http://amigo.geneontology.org/amigo/term/GO:0060370	susceptibility to T cell mediated cytotoxicity
http://amigo.geneontology.org/amigo/term/GO:0042110	T cell activation
http://amigo.geneontology.org/amigo/term/GO:0072678	T cell migration
http://amigo.geneontology.org/amigo/term/GO:0042098	T cell proliferation
http://amigo.geneontology.org/amigo/term/GO:0042608	T cell receptor binding

Vemos que obtenemos funciones moleculares y procesos biológicos donde intervienen proteínas en relación con las células T.

Consulta 6: haremos uso del operador “Union” para buscar cualquier entidad que tenga como propiedad ser de tipo molécula y que tenga un comentario y/o una etiqueta. Además, usamos “Order by” para pedirlos en orden alfabético.

```

1 prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
2 prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 prefix uniprot:<http://purl.uniprot.org/uniprot/>
4 prefix go: <http://amigo.geneontology.org/amigo/term/>
5 prefix uniprott:<http://purl.uniprot.org/core/>
6 prefix up:<https://www.uniprot.org/core/>
7 prefix gen:<https://www.ncbi.nlm.nih.gov/gene/>
8 prefix dis:<https://www.disease-ontology.org/do/>
9 prefix far:<https://go.drugbank.com/drugs/>
10 prefix onto:<http://purl.obolibrary.org/obo/>
11 prefix pub:<https://pubmed.ncbi.nlm.nih.gov/>
12 prefix var:<https://www.ebi.ac.uk/gwas/variants/>
13
14 SELECT ?p ?n
15 WHERE {
16   {?p rdf:type uniprott:Molecule.
17    ?p rdfs:comment ?n}
18 UNION
19   {?p rdf:type uniprott:Molecule.
20    ?p rdfs:label ?n}
21 }
22 ORDER BY ?n

```

P	n
< http://purl.uniprot.org/uniprot/P42081 >	CD86_HUMAN
< https://go.drugbank.com/drugs/DB05092 >	CDP323
< http://purl.uniprot.org/uniprot/P01911 >	DRB1_HUMAN
< https://go.drugbank.com/drugs/DB06637 >	Dalfampridine
< https://go.drugbank.com/drugs/DB08908 >	Dimethyl fumarate
< https://go.drugbank.com/drugs/DB14783 >	Diroximel fumarate
< https://go.drugbank.com/drugs/DB08868 >	Fingolimod
< https://go.drugbank.com/drugs/DB05259 >	Glatiramer
< http://purl.uniprot.org/uniprot/P01911 >	HLA class II histocompatibility antigen, DRB1 beta chain
< http://purl.uniprot.org/uniprot/P17181 >	INAR1_HUMAN
< http://purl.uniprot.org/uniprot/P13612 >	ITA4_HUMAN
< http://purl.uniprot.org/uniprot/P13612 >	Integrin alpha-4
< http://purl.uniprot.org/uniprot/P17181 >	Interferon alpha/beta receptor 1
< http://purl.uniprot.org/uniprot/P02686 >	MBP_HUMAN
< https://go.drugbank.com/drugs/DB01204 >	Mitoxantrone
< http://purl.uniprot.org/uniprot/P02686 >	Myelin basic protein
< https://go.drugbank.com/drugs/DB12612 >	Ozanimod
< https://go.drugbank.com/drugs/DB12016 >	Ponesimod
< http://purl.uniprot.org/uniprot/P21453 >	S1PR1_HUMAN
< https://go.drugbank.com/drugs/DB12371 >	Siponimod
< http://purl.uniprot.org/uniprot/P21453 >	Sphingosine 1-phosphate receptor 1
< http://purl.uniprot.org/uniprot/P42081 >	T-lymphocyte activation antigen CD86
< https://go.drugbank.com/drugs/DB08880 >	Teriflunomide

Consulta 7: haremos uso del operador “**Ask**” para preguntar si tres variedades concretas están asociadas con un incremento de la severidad en la enfermedad E.múltiple. Nos devuelve valor true, por lo que si están asociadas en nuestro dataset.

```

1 prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
2 prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 prefix uniprot:<http://purl.uniprot.org/uniprot/>
4 prefix go: <http://amigo.geneontology.org/amigo/term/>
5 prefix uniprott:<http://purl.uniprot.org/core/>
6 prefix up:<https://www.uniprot.org/core/>
7 prefix gen:<https://www.ncbi.nlm.nih.gov/gene/>
8 prefix dis:<https://www.disease-ontology.org/do/>
9 prefix far:<https://go.drugbank.com/drugs/>
10 prefix onto:<http://purl.obolibrary.org/obo/>
11 prefix pub:<https://pubmed.ncbi.nlm.nih.gov/>
12 prefix var:<https://www.ebi.ac.uk/gwas/variants/>
13
14 ASK
15 WHERE {
16
17   var:rs1478091-C onto:RO_0003305 dis:DOID:2377.
18   var:rs7191888-G onto:RO_0003305 dis:DOID:2377.
19   var:rs12638253-T onto:RO_0003305 dis:DOID:2377.
20 }
```

[Advanced features](#)

Execute Clear

true

Consulta 8: haremos uso del operador “**Describe**” para solicitar una descripción de la estructura RDF correspondiente a una proteína concreta, en este caso P13612.

```

1 prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
2 prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 prefix uniprot:<http://purl.uniprot.org/uniprot/>
4 prefix go: <http://amigo.geneontology.org/amigo/term/>
5 prefix uniprott:<http://purl.uniprot.org/core/>
6 prefix up:<https://www.uniprot.org/core/>
7 prefix gen:<https://www.ncbi.nlm.nih.gov/gene/>
8 prefix dis:<https://www.disease-ontology.org/do/>
9 prefix far:<https://go.drugbank.com/drugs/>
10 prefix onto:<http://purl.obolibrary.org/obo/>
11 prefix pub:<https://pubmed.ncbi.nlm.nih.gov/>
12 prefix var:<https://www.ebi.ac.uk/gwas/variants/>
13
14 DESCRIBE uniprot:P13612
```

subject	predicate	object
<http://purl.uniprot.org/uniprot/P13612>	<http://purl.obolibrary.org/obo/RO_0000056>	<http://amigo.geneontology.org/amigo/term/GO:0005178>
<http://purl.uniprot.org/uniprot/P13612>	<http://purl.obolibrary.org/obo/RO_0000056>	<http://amigo.geneontology.org/amigo/term/GO:0050839>
<http://purl.uniprot.org/uniprot/P13612>	<http://purl.obolibrary.org/obo/RO_0000056>	<http://amigo.geneontology.org/amigo/term/GO:1990138>
<http://purl.uniprot.org/uniprot/P13612>	<http://purl.obolibrary.org/obo/RO_0000056>	<http://amigo.geneontology.org/amigo/term/GO:2000406>
<http://purl.uniprot.org/uniprot/P13612>	<http://purl.uniprot.org/uniprot/citation>	<https://pubmed.ncbi.nlm.nih.gov/18635536>
<http://purl.uniprot.org/uniprot/P13612>	<http://purl.uniprot.org/uniprot/citation>	<https://pubmed.ncbi.nlm.nih.gov/21873635>
<http://purl.uniprot.org/uniprot/P13612>	<http://purl.uniprot.org/uniprot/encodedBy>	<https://www.ncbi.nlm.nih.gov/gene/3676>
<http://purl.uniprot.org/uniprot/P13612>	<https://www.uniprot.org/core/interaction>	<http://purl.uniprot.org/uniprot/Q86YTR>
<http://purl.uniprot.org/uniprot/P13612>	<https://www.uniprot.org/core/interaction>	<https://go.drugbank.com/drugs/DB00108>
<http://purl.uniprot.org/uniprot/P13612>	<https://www.uniprot.org/core/interaction>	<https://go.drugbank.com/drugs/DB05092>
<http://purl.uniprot.org/uniprot/P13612>	<https://www.uniprot.org/core/locatedIn>	<http://amigo.geneontology.org/amigo/term/GO:0034668>
<http://purl.uniprot.org/uniprot/P13612>	<https://www.uniprot.org/core/locatedIn>	<http://amigo.geneontology.org/amigo/term/GO:0043025>
<http://purl.uniprot.org/uniprot/P13612>	rdf:type	<http://purl.uniprot.org/core/Molecule>
<http://purl.uniprot.org/uniprot/P13612>	rdf:type	<http://purl.uniprot.org/core/Protein>
<http://purl.uniprot.org/uniprot/P13612>	rdfs:comment	ITAA4_HUMAN
<http://purl.uniprot.org/uniprot/P13612>	rdfs:label	Integrin alpha-4

5. Consultas SPARQL desde RStudio

El próximo paso será crear un script en RStudio desde el servidor del máster que nos permita realizar consultas externas en nuestro dataset cargado en Blazegraph, por medio de nuestra dirección IP vinculada (35.225.22.162) y el puerto utilizado (3031).

Realizaremos 5 de las consultas anteriores de esta manera y para ello primeramente necesitaremos cargar los siguientes paquetes y librerías.

```

1 install.packages("XML")
2 install.packages("RCurl")
3 library(XML)
4 library(RCurl)
5 library(SPARQL)

```

Consulta 1: devuelve todas las variantes génicas en secuencia, su etiqueta y los genes que la contienen.

```

7 #Consulta 1. Consulta genes y sus variantes y devuelve la etiqueta de las variantes
8 endpoint <- "http://35.225.22.162:3031/blazegraph/sparql"
9 query1 <-
10 prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
11 prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
12 prefix uniprot:<http://purl.uniprot.org/uniprot/>
13 prefix go: <http://amigo.geneontology.org/amigo/term/>
14 prefix uniprott:<http://purl.uniprot.org/core/>
15 prefix up:<https://www.uniprot.org/core/>
16 prefix gen:<https://www.ncbi.nlm.nih.gov/gene/>
17 prefix dis:<https://www.disease-ontology.org/do/>
18 prefix far:<https://go.drugbank.com/drugs/>
19 prefix onto:<http://purl.obolibrary.org/obo/>
20 prefix pub:<https://pubmed.ncbi.nlm.nih.gov/>
21 prefix var:<https://www.ebi.ac.uk/gwas/variants/>
22 SELECT ?var ?gen ?label
23 WHERE {
24   ?var onto:PORO_0000946 ?gen.
25   ?var rdfs:label ?label
26 }
27 "
28
29 qd1<- SPARQL(endpoint,query1)
30 View((qd1$results))

```

var	gen	label
1 <https://www.ebi.ac.uk/gwas/variants/rs10243024-A>	<https://www.ncbi.nlm.nih.gov/gene/4233>	variant rs10243024-A
2 <https://www.ebi.ac.uk/gwas/variants/rs10259085-C>	<https://www.ncbi.nlm.nih.gov/gene/56913>	variant rs10259085-C
3 <https://www.ebi.ac.uk/gwas/variants/rs12638253-T>	<https://www.ncbi.nlm.nih.gov/gene/389170>	variant rs12638253-T
4 <https://www.ebi.ac.uk/gwas/variants/rs1478091-C>	<https://www.ncbi.nlm.nih.gov/gene/100421232>	variant rs1478091-C
5 <https://www.ebi.ac.uk/gwas/variants/rs1478091-C>	<https://www.ncbi.nlm.nih.gov/gene/106479666>	variant rs1478091-C
6 <https://www.ebi.ac.uk/gwas/variants/rs180358-T>	<https://www.ncbi.nlm.nih.gov/gene/101929011>	variant rs180358-T
7 <https://www.ebi.ac.uk/gwas/variants/rs180358-T>	<https://www.ncbi.nlm.nih.gov/gene/84811>	variant rs180358-T
8 <https://www.ebi.ac.uk/gwas/variants/rs4953911-T>	<https://www.ncbi.nlm.nih.gov/gene/4249>	variant rs4953911-T
9 <https://www.ebi.ac.uk/gwas/variants/rs7191888-G>	<https://www.ncbi.nlm.nih.gov/gene/463>	variant rs7191888-G

Consulta 2: pediremos cualquier entidad que tenga un comentario y/o una etiqueta y queremos que guarde relación con las células T.

```

34 #Consulta 2. Consulto propiedades que tengan una etiqueta y/o comentario relacionado
35 #con linfocitos T (implicados en la respuesta autoinmune)
36 endpoint <- "http://35.225.22.162:3031/blazegraph/sparql"
37 query2 <-
38 prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
39 prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
40 prefix uniprot:<http://purl.uniprot.org/uniprot/>
41 prefix go: <http://amigo.geneontology.org/amigo/term/>
42 prefix uniprot:<http://purl.uniprot.org/core/>
43 prefix up:<https://www.uniprot.org/core/>
44 prefix onto:<http://purl.obolibrary.org/obo/>
45 prefix pub:<https://pubmed.ncbi.nlm.nih.gov/>
46 SELECT ?p ?n
47 WHERE {
48   {?p rdfs:label ?n}
49   UNION
50   {?p rdfs:comment ?n}
51   FILTER regex (?n, 'T cell.*')
52 }
53 "
54 qd2<- SPARQL(endpoint,query2)
55 View((qd2$results))

```

p	n
1 <http://amigo.geneontology.org/amigo/term/GO:2000406>	positive regulation of T cell migration
2 <http://amigo.geneontology.org/amigo/term/GO:0060370>	susceptibility to T cell mediated cytotoxicity
3 <http://amigo.geneontology.org/amigo/term/GO:0042110>	T cell activation
4 <http://amigo.geneontology.org/amigo/term/GO:0072678>	T cell migration
5 <http://amigo.geneontology.org/amigo/term/GO:0042098>	T cell proliferation
6 <http://amigo.geneontology.org/amigo/term/GO:0042608>	T cell receptor binding

Consulta 3: buscamos fármacos sean tratamiento de enfermedad (junto su etiqueta) y la enfermedad que tratan, usando el operador “Filter” para filtrar por tipo de enfermedad recurrente dentro de los subtipos de esclerosis múltiple registrados.

```

58 #Consulta 3. Consulta farmacos usados como tratamiento en E.multiple de tipo recurrente
59 #mostrando etiqueta tanto del farmaco como de la enfermedad
60 endpoint <- "http://35.225.22.162:3031/blazegraph/sparql"
61 query3 <-
62 prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
63 prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
64 prefix uniprot:<http://purl.uniprot.org/uniprot/>
65 prefix go: <http://amigo.geneontology.org/amigo/term/>
66 prefix uniprott:<http://purl.uniprot.org/core/>
67 prefix up:<https://www.uniprot.org/core/>
68 prefix dis:<https://www.disease-ontology.org/do/>
69 prefix far:<https://go.drugbank.com/drugs/>
70 prefix onto:<http://purl.obolibrary.org/obo/>
71 SELECT ?far ?label ?n
72 WHERE {
73   ?far onto:NCIT_C49236 ?dis.
74   ?far rdfs:label ?label.
75   ?dis rdfs:label ?n.
76   FILTER regex (?n, 'relapsing.*')
77 }
78 "
79
80 qd3<- SPARQL(endpoint,query3)
81 View((qd3$results))

```

	far	label	n
1	<https://go.drugbank.com/drugs/DB00108>	Natalizumab	progressive relapsing multiple sclerosis
2	<https://go.drugbank.com/drugs/DB01204>	Mitoxantrone	progressive relapsing multiple sclerosis
3	<https://go.drugbank.com/drugs/DB08880>	Teriflunomide	progressive relapsing multiple sclerosis
4	<https://go.drugbank.com/drugs/DB08908>	Dimethyl fumarate	progressive relapsing multiple sclerosis
5	<https://go.drugbank.com/drugs/DB12016>	Ponesimod	progressive relapsing multiple sclerosis
6	<https://go.drugbank.com/drugs/DB12371>	Siponimod	progressive relapsing multiple sclerosis
7	<https://go.drugbank.com/drugs/DB12612>	Ozanimod	progressive relapsing multiple sclerosis
8	<https://go.drugbank.com/drugs/DB14783>	Diroximel fumarate	progressive relapsing multiple sclerosis
9	<https://go.drugbank.com/drugs/DB14999>	Human interferon beta	progressive relapsing multiple sclerosis
10	<https://go.drugbank.com/drugs/DB00068>	recombinant human Interferon beta-1b	relapsing-remitting multiple sclerosis
11	<https://go.drugbank.com/drugs/DB00108>	Natalizumab	relapsing-remitting multiple sclerosis
12	<https://go.drugbank.com/drugs/DB01204>	Mitoxantrone	relapsing-remitting multiple sclerosis
13	<https://go.drugbank.com/drugs/DB05259>	Glatiramer	relapsing-remitting multiple sclerosis
14	<https://go.drugbank.com/drugs/DB08868>	Fingolimod	relapsing-remitting multiple sclerosis

Consulta 4: buscamos proteínas que se haya registrado el gen que las codifica, genes que tienen variantes de secuencia asociadas (pedimos esas variantes) y enfermedades que tienen variantes asociadas (pedimos esas variantes).

```

83 #Consulta 4. Consulta proteinas y gen por el que estan codificadas, variantes de los genes
84 #y las variantes asociadas a enfermedad
85 endpoint <- "http://35.225.22.162:3031/blazegraph/sparql"
86 query4 <-
87 prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
88 prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
89 prefix uniprot:<http://purl.uniprot.org/uniprot/>
90 prefix go: <http://amigo.geneontology.org/amigo/term/>
91 prefix uniprott:<http://purl.uniprot.org/core/>
92 prefix up:<https://www.uniprot.org/core/>
93 prefix gen:<https://www.ncbi.nlm.nih.gov/gene/>
94 prefix dis:<https://www.disease-ontology.org/do/>
95 prefix onto:<http://purl.obolibrary.org/obo/>
96 prefix var:<https://www.ebi.ac.uk/gwas/variants/>
97 SELECT ?protein ?gen ?var ?dis
98 WHERE {
99   ?protein uniprot:encodedBy ?gen.
100  ?gen onto:SO_0001060 ?var.
101  ?dis onto:SO_0001771 ?var
102 }
103 ..
104
105 qd4<- SPARQL(endpoint,query4)
106 View((qd4$results))

```

	protein	gen	var	dis
1	<http://purl.uniprot.org/uniprot/P21453>	<https://www.ncbi.nlm.nih.gov/gene/1901>	<https://www.ebi.ac.uk/gwas/variants/rs12756986-C>	<https://www.disease-ontology.org/do/DOID:2377>
2	<http://purl.uniprot.org/uniprot/P21453>	<https://www.ncbi.nlm.nih.gov/gene/1901>	<https://www.ebi.ac.uk/gwas/variants/rs17123757-C>	<https://www.disease-ontology.org/do/DOID:2377>
3	<http://purl.uniprot.org/uniprot/P42081>	<https://www.ncbi.nlm.nih.gov/gene/942>	<https://www.ebi.ac.uk/gwas/variants/rs2255214-C>	<https://www.disease-ontology.org/do/DOID:2377>
4	<http://purl.uniprot.org/uniprot/P42081>	<https://www.ncbi.nlm.nih.gov/gene/942>	<https://www.ebi.ac.uk/gwas/variants/rs2255214-T>	<https://www.disease-ontology.org/do/DOID:2377>
5	<http://purl.uniprot.org/uniprot/P42081>	<https://www.ncbi.nlm.nih.gov/gene/942>	<https://www.ebi.ac.uk/gwas/variants/rs2681424-C>	<https://www.disease-ontology.org/do/DOID:2377>
6	<http://purl.uniprot.org/uniprot/P42081>	<https://www.ncbi.nlm.nih.gov/gene/942>	<https://www.ebi.ac.uk/gwas/variants/rs2681424-T>	<https://www.disease-ontology.org/do/DOID:2377>
7	<http://purl.uniprot.org/uniprot/P42081>	<https://www.ncbi.nlm.nih.gov/gene/942>	<https://www.ebi.ac.uk/gwas/variants/rs4308217-C>	<https://www.disease-ontology.org/do/DOID:2377>
8	<http://purl.uniprot.org/uniprot/P17181>	<https://www.ncbi.nlm.nih.gov/gene/3454>	<https://www.ebi.ac.uk/gwas/variants/rs71314176-G>	<https://www.disease-ontology.org/do/DOID:2377>
9	<http://purl.uniprot.org/uniprot/P01911>	<https://www.ncbi.nlm.nih.gov/gene/3123>	<https://www.ebi.ac.uk/gwas/variants/rs9271366-G>	<https://www.disease-ontology.org/do/DOID:2377>
10	<http://purl.uniprot.org/uniprot/P01911>	<https://www.ncbi.nlm.nih.gov/gene/3123>	<https://www.ebi.ac.uk/gwas/variants/rs9271640-T>	<https://www.disease-ontology.org/do/DOID:2377>
11	<http://purl.uniprot.org/uniprot/P42081>	<https://www.ncbi.nlm.nih.gov/gene/942>	<https://www.ebi.ac.uk/gwas/variants/rs9282641-G>	<https://www.disease-ontology.org/do/DOID:2377>

Consulta 5: buscar cualquier entidad que tenga como propiedad ser de tipo molécula y que tenga un comentario y/o una etiqueta. Además, pedimos en orden alfabético.

```

109 #Consulta 5. Consulta todas las propiedades que sean moléculas y que tengan una etiqueta y/o
110 #un comentario ordenados alfabéticamente
111 endpoint <- "http://35.225.22.162:3031/blazegraph/sparql"
112 query5 <-
113 prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
114 prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
115 prefix uniprot:<http://purl.uniprot.org/uniprot/>
116 prefix go: <http://amigo.geneontology.org/amigo/term/>
117 prefix uniprott:<http://purl.uniprot.org/core/>
118 prefix up:<https://www.uniprot.org/core/>
119 prefix far:<https://go.drugbank.com/drugs/>
120 prefix onto:<http://purl.obolibrary.org/obo/>
121 SELECT ?p ?n
122 WHERE {
123   {?p rdf:type uniprott:Molecule.
124    ?p rdfs:comment ?n}
125   UNION
126   {?p rdf:type uniprott:Molecule.
127    ?p rdfs:label ?n}
128 }
129 ORDER BY ?n
130 "
131
132 qd5<- SPARQL(endpoint,query5)
133 View((qd5$results))

```

	p	n
1	<http://purl.uniprot.org/uniprot/P42081>	CD86_HUMAN
2	<https://go.drugbank.com/drugs/DB05092>	CDP323
3	<http://purl.uniprot.org/uniprot/P01911>	DRB1_HUMAN
4	<https://go.drugbank.com/drugs/DB06637>	Dalfampridine
5	<https://go.drugbank.com/drugs/DB08908>	Dimethyl fumarate
6	<https://go.drugbank.com/drugs/DB14783>	Diroximel fumarate
7	<https://go.drugbank.com/drugs/DB08868>	Fingolimod
8	<https://go.drugbank.com/drugs/DB05259>	Glatiramer
9	<http://purl.uniprot.org/uniprot/P01911>	HLA class II histocompatibility antigen, DRB1 beta chain
10	<http://purl.uniprot.org/uniprot/P17181>	INAR1_HUMAN
11	<http://purl.uniprot.org/uniprot/P13612>	ITA4_HUMAN
12	<http://purl.uniprot.org/uniprot/P13612>	Integrin alpha-4
13	<http://purl.uniprot.org/uniprot/P17181>	Interferon alpha/beta receptor 1
14	<http://purl.uniprot.org/uniprot/P02686>	MBP_HUMAN

Showing 1 to 15 of 23 entries, 2 total columns

6. Publicación de los datos y aplicación de principios FAIR

El siguiente paso será publicar nuestro dataset siguiendo los principios FAIR. Para proceder a su publicación en Blazegraph, en primer lugar, debemos crear un namespace para poder representar nuestro dataset y los metadatos que lo definen juntos en RDF. Nosotros vamos a escoger el nombre de namespace “**MultipleSclerosisPathway**” en modo quads y una vez creado lo pondremos en uso.

The screenshot shows the Blazegraph Workbench interface. At the top, there's a logo and the text "blazegraph workbench ultra-scalable, high-performance database from Blazegraph". Below the header, there are tabs: WELCOME (selected), QUERY, UPDATE, EXPLORE, NAMESPACES, STATUS, and PERFORMANCE. The NAMESPACES tab is active, showing a table with two rows:

MultipleSclerosisPathway	In use	Delete	Properties	Rebuild Full Text Index	Clone	Service Description
kb	Use	Delete	Properties	Rebuild Full Text Index	Clone	Service Description

Below the table, there's a link "Download VOID description of all namespaces". Under the "Namespaces" heading, there's a section titled "Create namespace" with instructions: "There are a number of features to enable. There's full documentation [here](#). You must select "Use" after you have created the KB." It also says "A quick reference is below:" followed by three bullet points about PropertyGraph, RDF + SPARQL, and RIFCID. At the bottom of this section, there's a form with fields for "Name:" (containing "MultipleSclerosisPathway"), "Mode:" (set to "quads"), and "Inference:" (unchecked). A note states "(Inference disabled - inference is not compatible with quads or isolatable indices.) Isolatable ind". Finally, there's a "Create namespace" button.

Figura 8.- Creación de namespace en Blazegraph

Llegados a este punto, es el momento de conectar nuestro dataset con los metadatos que lo definen. Nosotros tenemos nuestros datos y metadatos ambos en formato Turtle (Terse RDF Triple Language) pero para poder conectarlos es necesario que nuestro dataset tenga una URI de contexto en la que están incluidos, que corresponderá a la URI namedGraph y que incluimos en nuestros metadatos. En este caso dicha URI es:<<https://um.es/graph/MultipleSclerosisDataset>>.

De una forma más concreta, lo que haremos será convertir nuestro dataset en Turtle a N-Triples usando la herramienta <http://www.easyrdf.org/converter> y con un pequeño script de Python reemplazaremos el punto final de cada tripleta por nuestra cuarta URI de contexto acabada en punto final, convirtiendo nuestro dataset en N-Quads (figura 10).

Sin embargo, es aquí cuando nos damos cuenta de que posteriormente cuando accedamos a nuestro dataset publicado desde Pubby, no podremos realizar consultas específicas ya que tenemos nuestro RDF con prefijos directos a las URI externas sin relacionar con una URI nuestra.

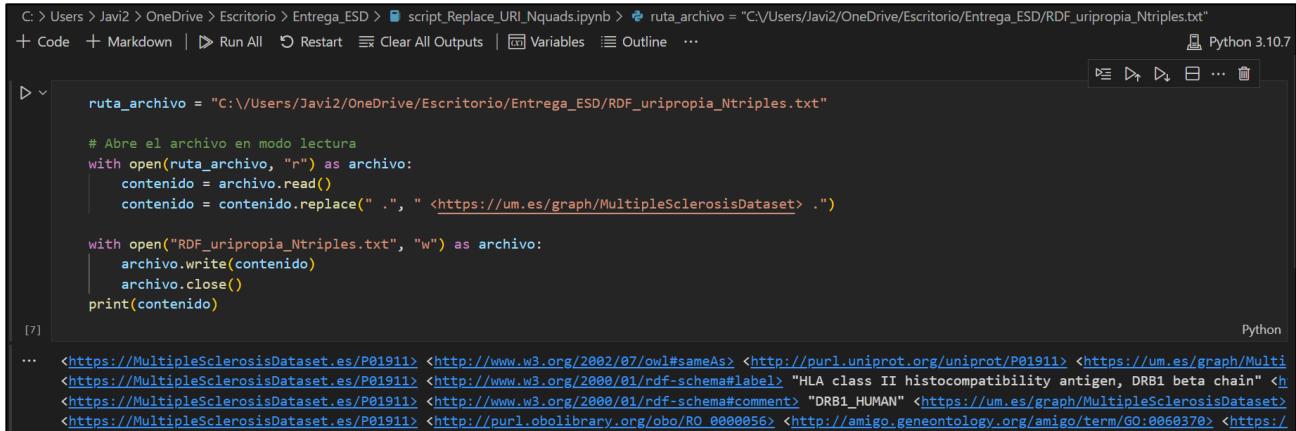
Para que sean posibles las consultas, antes de realizar el paso a N-Triples, debemos incluir una URI propia de nuestro dataset en todos los sujetos de nuestro RDF Turtle, que será la que incluiremos en el archivo de metadatos (<<https://MultipleSclerosisDataset.es/>>) y con la propiedad owl:sameAs le indicamos que nos rescate la información desde las URI externas de proteínas, genes, variantes, fármacos, enfermedades y ontologías.

```
@prefix em: <https://MultipleSclerosisDataset.es/>.
@prefix owl: <http://www.w3.org/2002/07/owl#>.

em:P01911
    owl:sameAs uniprot:P01911;
```

Figura 9.- Adición de URI propia de dataset a RDF Turtle (tripletas_entrega_URI_propia.txt)

A continuación, es el momento de convertir RDF Turtle a N-Triples y con Python a N-Quads de la siguiente manera:



```
C: > Users > Javi2 > OneDrive > Escritorio > Entrega_ESD > script_Replace_URL_Nquads.ipynb > ruta_archivo = "C:/Users/Javi2/OneDrive/Escritorio/Entrega_ESD/RDF_uripropia_Ntriples.txt"
+ Code + Markdown | ▶ Run All ⚡ Restart ⌂ Clear All Outputs | ⓘ Variables ⌂ Outline ... Python 3.10.7
[7] ruta_archivo = "C:/Users/Javi2/OneDrive/Escritorio/Entrega_ESD/RDF_uripropia_Ntriples.txt"

# Abre el archivo en modo lectura
with open(ruta_archivo, "r") as archivo:
    contenido = archivo.read()
    contenido = contenido.replace(" .", " <https://um.es/graph/MultipleSclerosisDataset> .")

with open("RDF_uripropia_Ntriples.txt", "w") as archivo:
    archivo.write(contenido)
    archivo.close()
print(contenido)
...
<https://MultipleSclerosisDataset.es/P01911> <http://www.w3.org/2002/07/owl#sameAs> <http://purl.uniprot.org/uniprot/P01911> <https://um.es/graph/Multi>
<https://MultipleSclerosisDataset.es/P01911> <http://www.w3.org/2000/01/rdf-schema#label> "HLA class II histocompatibility antigen, DRB1 beta chain" <https://um.es/graph/Multi>
<https://MultipleSclerosisDataset.es/P01911> <http://www.w3.org/2000/01/rdf-schema#comment> "DRB1_HUMAN" <https://um.es/graph/MultipleSclerosisDataset>
<https://MultipleSclerosisDataset.es/P01911> <http://purl.obolibrary.org/obo/R0\_0000056> <http://amigo.geneontology.org/amigo/term/GO:0060370> <https://um.es/graph/Multi>
```

Figura 10.- Script N-quads Python

```
<https://MultipleSclerosisDataset.es/P01911> <http://www.w3.org/2002/07/owl#sameAs> <http://purl.uniprot.org/uniprot/P01911>
<https://um.es/graph/MultipleSclerosisDataset> .
<https://MultipleSclerosisDataset.es/P01911> <http://www.w3.org/2000/01/rdf-schema#label> "HLA class II histocompatibility antigen, DRB1 beta chain" <https://um.es/graph/Multi>
<https://um.es/graph/MultipleSclerosisDataset> .
```

Figura 11.- Extracto archivo RDF_uripropia_NQuads.txt

Evidentemente, nuestro RDF Turtle ahora será distinto y tendrá más de 349 triplets y nuestro RDF N-Quads también cambiará en consecuencia teniendo más triplets.

A continuación, indicaremos en nuestro archivo de metadatos esa URI de contexto como namedGraph. Además indicamos cual será nuestro endpoint de acceso a Blazegraph, con el puerto especificado, una descripción y autor del grafo realizado en este caso.

```

@prefix dcat: <http://www.w3.org/ns/dcat#>.
@prefix dct: <http://purl.org/dc/terms/>.
@prefix sd: <http://www.w3.org/ns/sparql-service-description#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix foaf: <http://xmlns.com/foaf/0.1/>.
@prefix cc: <http://creativecommons.org/ns#>.

<https://MultipleSclerosisDataset.es/> rdf:type dcat:Dataset ;
    rdfs:label "Multiple Sclerosis Dataset" ;
    foaf:primaryTopic "Multiple Sclerosis Pathway" ;
    dct:License <http://creativecommons.org/licenses/MIT/> ;
    dcat:distribution <https://um.es/txt/MultipleSclerosisDataset> ;
    dcat:distribution <https://um.es/sparql/MultipleSclerosisDataset> ;
    sd:namedGraph <https://um.es/graph/MultipleSclerosisDataset> .

<https://um.es/txt/MultipleSclerosisDataset> a dcat:Distribution ;
    dcat:downloadURL <http://localhost/RDF_NQuads.txt> ;
    dct:license <http://creativecommons.org/licenses/by/3.0/es/> ;
    dct:format <http://publications.europa.eu/resource/authority/file-type/TXT> .

<https://um.es/sparql/MultipleSclerosisDataset> a dcat:Distribution ;
    dcat:accessURL <http://http://35.225.22.162:3031/blazegraph/namespace/MultipleSclerosisPathway/sparql> ;
    dct:license <http://creativecommons.org/licenses/by/3.0/es/> .

<https://um.es/graph/MultipleSclerosisDataset> rdfs:label "Grafo RDF sobre Esclerosis Multiple"@es ;
    rdfs:type sd:namedGraph ;
    dct:creator "F.Javier Lopez Carbonell"@es .

<http://creativecommons.org/licenses/MIT/>
    a cc:License ;
    cc:permits cc:Distribution, cc:Reproduction, cc:DerivativeWorks ;
    cc:licenseClass <http://creativecommons.org/license/software> ;

```

Figura 12.- Contenido del archivo de metadatos (archivo: MetadatosEMdataset.txt)

Una vez tenemos metadatos y RDF en N-Quads configurados, procedemos a actualizarlos en Blazegraph. Para ello nos dirigimos a la pestaña “update” y seleccionamos nuestro archivo RDF_uripropia_NQuads.txt en tipo RDF Data y en formato N-Quads. Ahora, seleccionamos nuestro archivo de metadatos MetadatosEMdataset.txt en tipo RDF Data y en formato Turtle, como se muestra en la figura 13.

The screenshot shows the Blazegraph update interface. The main area displays the RDF code from Figure 12. Below the code, there are dropdown menus for 'Seleccionar archivo' (Select file) and 'Type' (set to 'RDF Data') and 'Format' (set to 'Turtle'). At the bottom left, there is a link 'Advanced features'.

```

1 @prefix dcat: <http://www.w3.org/ns/dcat#>.
2 @prefix dct: <http://purl.org/dc/terms/>.
3 @prefix sd: <http://www.w3.org/ns/sparql-service-description#>.
4 @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
5 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
6 @prefix foaf: <http://xmlns.com/foaf/0.1/>.
7 @prefix cc: <http://creativecommons.org/ns#>.
8
9 <https://MultipleSclerosisDataset.es/> rdf:type dcat:Dataset ;
    rdfs:label "Multiple Sclerosis Dataset" ;
    foaf:primaryTopic "Multiple Sclerosis Pathway" ;
    dct:License <http://creativecommons.org/licenses/MIT/> ;
    dcat:distribution <https://um.es/txt/MultipleSclerosisDataset> ;
    dcat:distribution <https://um.es/sparql/MultipleSclerosisDataset> ;
    sd:namedGraph <https://um.es/graph/MultipleSclerosisDataset> .
10
11 <https://um.es/txt/MultipleSclerosisDataset> a dcat:Distribution ;
12     dcat:downloadURL <http://localhost/RDF_NQuads.txt> ;
13     dct:license <http://creativecommons.org/licenses/by/3.0/es/> ;
14     dct:format <http://publications.europa.eu/resource/authority/file-type/TXT> .
15
16 <https://um.es/sparql/MultipleSclerosisDataset> a dcat:Distribution ;
17     dcat:accessURL <http://http://35.225.22.162:3031/blazegraph/namespace/MultipleSclerosisPathway/sparql> ;
18
19 Seleccionar archivo Ninguno archivo selec.
20 Type: RDF Data Format: Turtle
21 Advanced features

```

Figura 13.- Archivo de metadatos subido en Blazegraph

Debemos comprobar si esta actualización de datos y metadatos en Blazegraph se ha realizado de forma correcta, por lo que realizamos una consulta de grafo (metadatos), sujetos, predicados y objetos, obteniendo un único conjunto de resultados mostrados en las figuras 14 y 15.

graph	s	p	
bd:nullGraph	http://creativecommons.org/licenses/MIT/	http://creativecommons.org/ns#legalcode	http://opensource.org/licenses/mit-lic
bd:nullGraph	http://creativecommons.org/licenses/MIT/	http://creativecommons.org/ns#licenseClass	http://creativecommons.org/license/soft
bd:nullGraph	http://creativecommons.org/licenses/MIT/	http://creativecommons.org/ns#permits	http://creativecommons.org/ns#Derivativ
bd:nullGraph	http://creativecommons.org/licenses/MIT/	http://creativecommons.org/ns#permits	http://creativecommons.org/ns#Distribut
bd:nullGraph	http://creativecommons.org/licenses/MIT/	http://creativecommons.org/ns#permits	http://creativecommons.org/ns#Reproduct
bd:nullGraph	http://creativecommons.org/licenses/MIT/	http://creativecommons.org/ns#resources	http://creativecommons.org/ns#seNotice
bd:nullGraph	http://creativecommons.org/licenses/MIT/	rdf:type	http://creativecommons.org/ns#License
bd:nullGraph	https://um.es/dataset/UMDataset/MultipleSclerosisDataset	dcterm:license	http://creativecommons.org/licenses/MIT
bd:nullGraph	https://um.es/dataset/UMDataset/MultipleSclerosisDataset	http://www.w3.org/ns/dc#distribution	https://um.es/sparql/MultipleSclerosisDataset
bd:nullGraph	https://um.es/dataset/UMDataset/MultipleSclerosisDataset	http://www.w3.org/ns/dc#distribution	https://um.es/txt/MultipleSclerosisDataset
bd:nullGraph	https://um.es/dataset/UMDataset/MultipleSclerosisDataset	http://www.w3.org/ns/sparql-service-description#namedGraph	https://um.es/graph/MultipleSclerosisDataset
bd:nullGraph	https://um.es/dataset/UMDataset/MultipleSclerosisDataset	rdf:type	http://www.w3.org/ns/dc#dataset
bd:nullGraph	https://um.es/dataset/UMDataset/MultipleSclerosisDataset	rdfs:label	Multiple Sclerosis Dataset
bd:nullGraph	https://um.es/graph/MultipleSclerosisDataset	dcterm:creator	F. Javier Lopez Carbonell
bd:nullGraph	https://um.es/graph/MultipleSclerosisDataset	rdf:type	http://www.w3.org/ns/sparql-service-des
bd:nullGraph	https://um.es/graph/MultipleSclerosisDataset	rdfs:label	Grafo RDF sobre Esclerosis Multiple
bd:nullGraph	https://um.es/sparql/MultipleSclerosisDataset	dcterm:license	http://creativecommons.org/licenses/by/
bd:nullGraph	https://um.es/sparql/MultipleSclerosisDataset	http://www.w3.org/ns/dc#accessURL	http://http://35.225.22.162:3031/blazeg
bd:nullGraph	https://um.es/sparql/MultipleSclerosisDataset	rdf:type	http://www.w3.org/ns/dc#Distribution
bd:nullGraph	https://um.es/txt/MultipleSclerosisDataset	dcterm:format	http://publications.europa.eu/resource/
bd:nullGraph	https://um.es/txt/MultipleSclerosisDataset	dcterm:license	http://creativecommons.org/licenses/by/
bd:nullGraph	https://um.es/txt/MultipleSclerosisDataset	http://www.w3.org/ns/dc#downloadURL	http://localhost/RDF_NQuads.txt
bd:nullGraph	https://um.es/txt/MultipleSclerosisDataset	rdf:type	http://www.w3.org/ns/dc#Distribution

Figura 14.- Consulta de metadatos en Blazegraph

https://um.es/graph/MultipleSclerosisDataset	http://amigo.geneontology.org/amigo/term/GO:0002020	rdfs:label	protease binding
https://um.es/graph/MultipleSclerosisDataset	http://amigo.geneontology.org/amigo/term/GO:0003376	rdfs:label	sphingosine-1-phosphate receptor signali
https://um.es/graph/MultipleSclerosisDataset	http://amigo.geneontology.org/amigo/term/GO:0003376	rdfs:subClassof	http://amigo.geneontology.org/amigo/ter
https://um.es/graph/MultipleSclerosisDataset	http://amigo.geneontology.org/amigo/term/GO:0004905	rdfs:label	type I Interferon receptor activity
https://um.es/graph/MultipleSclerosisDataset	http://amigo.geneontology.org/amigo/term/GO:0004905	rdfs:subClassof	http://amigo.geneontology.org/amigo/ter
https://um.es/graph/MultipleSclerosisDataset	http://amigo.geneontology.org/amigo/term/GO:0004930	rdfs:label	G protein-coupled receptor activity
https://um.es/graph/MultipleSclerosisDataset	http://amigo.geneontology.org/amigo/term/GO:0004930	rdfs:subClassof	http://amigo.geneontology.org/amigo/ter
https://um.es/graph/MultipleSclerosisDataset	http://amigo.geneontology.org/amigo/term/GO:0004930	rdfs:label	integrin binding
https://um.es/graph/MultipleSclerosisDataset	http://amigo.geneontology.org/amigo/term/GO:0005178	rdfs:subClassof	http://amigo.geneontology.org/amigo/ter
https://um.es/graph/MultipleSclerosisDataset	http://amigo.geneontology.org/amigo/term/GO:0005178	rdfs:label	immune response
https://um.es/graph/MultipleSclerosisDataset	http://amigo.geneontology.org/amigo/term/GO:0006955	rdfs:subClassof	http://amigo.geneontology.org/amigo/ter
https://um.es/graph/MultipleSclerosisDataset	http://amigo.geneontology.org/amigo/term/GO:0006955	rdfs:label	receptor signaling pathway via JAK-STAT
https://um.es/graph/MultipleSclerosisDataset	http://amigo.geneontology.org/amigo/term/GO:0007259	rdfs:label	http://amigo.geneontology.org/amigo/ter
https://um.es/graph/MultipleSclerosisDataset	http://amigo.geneontology.org/amigo/term/GO:0007259	rdfs:subClassof	central nervous system development
https://um.es/graph/MultipleSclerosisDataset	http://amigo.geneontology.org/amigo/term/GO:0007417	rdfs:label	http://amigo.geneontology.org/amigo/ter
https://um.es/graph/MultipleSclerosisDataset	http://amigo.geneontology.org/amigo/term/GO:0007417	rdfs:subClassof	JAK pathway signal transduction adaptor
https://um.es/graph/MultipleSclerosisDataset	http://amigo.geneontology.org/amigo/term/GO:0008269	rdfs:label	http://amigo.geneontology.org/amigo/ter
https://um.es/graph/MultipleSclerosisDataset	http://amigo.geneontology.org/amigo/term/GO:0008269	rdfs:subClassof	coreceptor activity
https://um.es/graph/MultipleSclerosisDataset	http://amigo.geneontology.org/amigo/term/GO:0015026	rdfs:label	http://amigo.geneontology.org/amigo/ter
https://um.es/graph/MultipleSclerosisDataset	http://amigo.geneontology.org/amigo/term/GO:0015026	rdfs:subClassof	structural constituent of myelin sheath
https://um.es/graph/MultipleSclerosisDataset	http://amigo.geneontology.org/amigo/term/GO:0019911	rdfs:label	http://amigo.geneontology.org/amigo/ter
https://um.es/graph/MultipleSclerosisDataset	http://amigo.geneontology.org/amigo/term/GO:0019911	rdfs:subClassof	MHC class II protein complex binding
https://um.es/graph/MultipleSclerosisDataset	http://amigo.geneontology.org/amigo/term/GO:0023026	rdfs:label	http://amigo.geneontology.org/amigo/ter
https://um.es/graph/MultipleSclerosisDataset	http://amigo.geneontology.org/amigo/term/GO:0023026	rdfs:subClassof	cellular response to interferon-beta

Figura 15.- Consulta de datos en Blazegraph

La consulta realizada para obtener el resultado de las figuras anteriores se muestra en la figura 16 a continuación.

```

1 SELECT ?graph ?s ?p ?o
2 WHERE {
3   GRAPH ?graph {
4     ?s ?p ?o
5   }
6 }
```

Figura 16.- Script consulta datos y metadatos en Blazegraph

Como podemos apreciar en las figuras 14 y 15 marcado en rojo, existe una conexión entre datos y metadatos de forma correcta por medio del namedGraph, aunque el resultado nos indique que el grafo es nulo.

Una vez publicados los datos en Blazegraph podemos hacer consultas directamente desde nuestro blazegraph, a través de diversas herramientas que posee el directorio LinkedDataServer clonado en nuestra instancia VM (miservidor) en GoogleCloud. Estas herramientas permiten negociar los datos y seguir aplicando los **principios FAIR** de publicación de datos:

- **Findable**
- **Accesible**
- **Interoperable**
- **Reusable**

La herramienta que nosotros vamos a utilizar es Pubby, que permite explorar datos RDF y descubrir nuevos datos a través de predicados. Debemos configurar Pubby previamente desde una ventana SSH distinta a la que usamos para ejecutar Blazegraph, de la siguiente manera:

```
franciscojavierlopezcarbonell@miservidor:~/UM-Bioinformatics-MSc-FAIR-data/LinkedDataServer/pubby$ ls
README.txt bin lib logs resources start.d start.jar
VERSION.txt etc license-eplv10-aslv20.html notice.html run.sh start.ini webapps
franciscojavierlopezcarbonell@miservidor:~/UM-Bioinformatics-MSc-FAIR-data/LinkedDataServer/pubby$ cd webapps/RO
OT/WEB-INF/
franciscojavierlopezcarbonell@miservidor:~/UM-Bioinformatics-MSc-FAIR-data/LinkedDataServer/pubby/webapps/ROOT/W
EB-INF$ ls
blazegraph-config.ttl classes config.ttl lib templates
books-config-file.ttl config-myfoaf.ttl fair-config.ttl lsld-toy-config-file.ttl web.xml
franciscojavierlopezcarbonell@miservidor:~/UM-Bioinformatics-MSc-FAIR-data/LinkedDataServer/pubby/webapps/ROOT/W
EB-INF$ nano fair-config.ttl
```

Figura 16.- Configuración Pubby en SSH

Con el uso del editor nano modificamos el archivo “**fair-config.ttl**” añadiendo nuestra IP y puerto seleccionado (3031) en “sparqlendpoint” de Blazegraph, nuestra URI propia del dataset y nuestra IP y nuestro puerto seleccionado para Pubby (8181). A continuación, este archivo fair-config.ttl debemos especificarlo en el archivo web.xml que se encuentra en una ruta de directorios especificada en el github de Mikel Egana junto con los pasos seguidos anteriormente.

Mencionar que para que la ejecución de Pubby funcione, antes debemos crear una regla de firewall para el puerto 8181 y debemos tener Blazegraph ejecutándose en una ventana SSH alternativa mientras ejecutamos Pubby. Ejecutamos Pubby de la siguiente manera

```
franciscojavierlopezcarbonell@miservidor:~$ cd UM-Bioinformatics-MSc-FAIR-data/LinkedDataServer/pubby/
franciscojavierlopezcarbonell@miservidor:~/UM-Bioinformatics-MSc-FAIR-data/LinkedDataServer/pubby$ java -jar sta
rt.jar jetty.port=8181
```

Figura 17.- Ejecución Pubby en SSH

Para comprobar si es correcto, accedemos a nuestro navegador y ponemos en el buscador nuestra IP number junto con el puerto 8181, en mi caso: <http://35.225.22.162:8181/> y este será el resultado:

Property	Value
?License	▪ <http://creativecommons.org/licenses/MIT/>
?distribution	▪ <https://um.es/sparql/MultipleSclerosisDataset> ▪ <https://um.es/xtl/MultipleSclerosisDataset>
?label	▪ Multiple Sclerosis Dataset ()
?namedGraph	▪ <https://um.es/graph/MultipleSclerosisDataset>
?primaryTopic	▪ Multiple Sclerosis Pathway ()
?type	▪ <http://www.w3.org/ns/dcat#Dataset>

This page shows information obtained from the SPARQL endpoint at <http://35.225.22.162:8181/blazegraph/namespaces/MultipleSclerosisPathway/sparql>.
[As.Turtle](#) | [As.RDF/XML](#) | [Browse in Disco](#) | [Browse in Tabulator](#) | [Browse in OpenLink Browser](#)

Como podemos comprobar, tenemos acceso a la estructura de nuestro grafo con las características especificadas por los metadatos actualizado en Blazegraph.

Sin embargo, podemos acceder a datos de nuestro data set de forma más específica y será parecido a realizar una consulta SPARQL con el operador “Describe” en Blazegraph.

Para ello, simplemente especificados la ID del dato tras el puerto 8181 en el navegador y aparecerá como se muestra a continuación con un ejemplo de proteína y de fármaco.

Property	Value
? RO_0000056	<ul style="list-style-type: none"> ▪ <http://amigo.geneontology.org/amigo/term/GO_0023026> ▪ <http://amigo.geneontology.org/amigo/term/GO_0042608> ▪ <http://amigo.geneontology.org/amigo/term/GO_0060370>
? citation	<ul style="list-style-type: none"> ▪ <https://pubmed.ncbi.nlm.nih.gov/19303388> ▪ <https://pubmed.ncbi.nlm.nih.gov/21833088>
? comment	<ul style="list-style-type: none"> ▪ DRB1_HUMAN ()
? encodedBy	<ul style="list-style-type: none"> ▪ <https://www.ncbi.nlm.nih.gov/gene/3123>
? interaction	<ul style="list-style-type: none"> ▪ <http://purl.uniprot.org/uniprot/P01903> ▪ <https://go.drugbank.com/drugs/DB05259>
? label	<ul style="list-style-type: none"> ▪ HLA class II histocompatibility antigen, DRB1 beta chain ()
? locatedIn	<ul style="list-style-type: none"> ▪ <http://amigo.geneontology.org/amigo/term/GO_0009986>
? sameAs	<ul style="list-style-type: none"> ▪ <http://purl.uniprot.org/uniprot/P01911> ▪ <http://purl.uniprot.org/core/Molecule> ▪ <http://purl.uniprot.org/core/Protein>
? type	<ul style="list-style-type: none"> ▪ <http://purl.uniprot.org/core/Molecule>

This page shows information obtained from the SPARQL endpoint at <http://35.225.22.162:3031/blazegraph/namespaces/MultipleSclerosisPathway/sparql>.
[As Turtle](#) | [As RDF/XML](#) | [Browse in Disco](#) | [Browse in Tabulator](#) | [Browse in OpenLink Browser](#)

Property	Value
? NCIT_C49236	<ul style="list-style-type: none"> ▪ <https://www.disease-ontology.org/do/DOID_0050783> ▪ <https://www.disease-ontology.org/do/DOID_0050784> ▪ <https://www.disease-ontology.org/do/DOID_0050785> ▪ <https://www.disease-ontology.org/do/DOID_2377> ▪ <https://www.disease-ontology.org/do/DOID_2378>
? label	<ul style="list-style-type: none"> ▪ Ozanimod ()
? sameAs	<ul style="list-style-type: none"> ▪ <https://go.drugbank.com/drugs/DB12612>
? type	<ul style="list-style-type: none"> ▪ <http://purl.obolibrary.org/obo/NCIT_C459> ▪ <http://purl.uniprot.org/core/Molecule>

This page shows information obtained from the SPARQL endpoint at <http://35.225.22.162:3031/blazegraph/namespaces/MultipleSclerosisPathway/sparql>.
[As Turtle](#) | [As RDF/XML](#) | [Browse in Disco](#) | [Browse in Tabulator](#) | [Browse in OpenLink Browser](#)

Recordar que, esto es posible gracias a que hemos especificado la URI propia de nuestro dataset en cada uno de los sujetos de nuestro script RDF Turtle y posteriormente en formato N-Quads subido a Blazegraph.

En caso de no hacerlo, la dirección que ponemos en "Dataset" en el archivo fair-config.ttl actúa como prefijo, pero no estaría especificado en nuestros datos.

Por tanto, al acceder desde Pubby no me va a dar como resultado los datos correspondientes a las URI externas de proteínas, fármacos, enfermedades, etc. Obtendremos en consecuencia el siguiente resultado:

404 Not Found at FAIR data publication UM MSc Bioinformatics
<http://35.225.22.162:8181/P21453>

The requested resource does not exist at this server, or no information about it is available.

This page shows information obtained from the SPARQL endpoint at <http://35.225.22.162:3031/blazegraph/namespaces/MultipleSclerosisPathway/sparql>.

Por último y a modo de conclusión, mencionar que hemos cumplido el objetivo de crear un conjunto de datos sobre la enfermedad esclerosis múltiple, reciclando datos y metadatos de bases de datos externas en lenguaje RDF, que es idóneo y el más recomendado para recopilar, exponer, consultar y compartir información por medio de diferentes aplicaciones o herramientas. Hemos sido capaces de cargar estos datos en el repositorio blazegraph y los hemos consultado tanto allí como en RStudio con el lenguaje de consulta SPARQL. Finalmente, hemos realizado una correcta publicación de los mismos siguiendo los principios FAIR.

7. Directorio de archivos generados

Archivos generados incluidos en la siguiente ruta de directorios de dayhoff.inf.um.es:

/home/alumno15/explotacion_semantica_datos/Practica_entregable

Destacar que he subido tantos los archivos RDF turtle y RDF N-Quads sin la URI propia de mi dataset como con dicha URI formando parte de las triplets.