

DATASET ANTICUERPOS COMERCIALES PARA EXPERIMENTACIÓN

EXPLOTACIÓN SEMÁNTICA DE DATOS

Curso 2023/2024

Fernando Lucas Ruiz

Tabla de contenido

OBJETIVO 3

CREACIÓN ONTOLOGÍA 4

CREACIÓN DE LOS INDIVIDUOS 5

VALIDACIÓN DEL DATASET 6

CONSULTAS SPARQL..... 7

BÚSQUEDAS DESDE EL PAQUETE SPARQL DE R..... 12

PUBLICACIÓN DEL DATASET..... 16

OBJETIVO

El objetivo principal de este proyecto consiste en crear un conjunto de datos RDF que contenga información sobre anticuerpos comerciales destinados a ensayos experimentales. Con este conjunto de datos, buscamos organizar una lista de anticuerpos y vincularla con bases de datos externas, facilitando a los investigadores la búsqueda de la información necesaria para sus ensayos.

La motivación detrás de la creación de este conjunto de datos radica en la frecuente utilización de anticuerpos de proveedores comerciales en experimentos de biología molecular. Estos anticuerpos suelen tener costos elevados, y en muchas ocasiones, los investigadores solo requieren pequeñas cantidades para llevar a cabo experimentos específicos, lo que resulta en la subutilización de viales completos de anticuerpos. Además, la amplia gama de opciones de anticuerpos disponibles para el reconocimiento de antígenos dificulta la elección del anticuerpo más adecuado para un ensayo particular.

La recopilación de datos sobre los anticuerpos adquiridos por diversos investigadores en institutos de investigación o universidades permite a los usuarios conocer la disponibilidad de un anticuerpo específico en su centro específico para realizar ensayos de prueba de concepto, obtener respuestas de revisores, evaluar la eficacia del anticuerpo en modelos específicos, entre otros usos. Esta iniciativa contribuye significativamente a la optimización de presupuestos en laboratorios y ahorro de espacio en congeladores.

Para llevar a cabo esta tarea, se desarrolló una ontología específica, diseñando clases que se ajustaran mejor a la información necesaria para esta tarea. Posteriormente, se incorporaron datos de 20 anticuerpos comerciales, estableciendo conexiones con la ontología previamente creada.

Nota: todos los documentos de esta tarea se encuentran en la carpeta **[/home/alumno14/ESD](#)**

La tabla con los 20 anticuerpos están situados en **[ESD/anticuerpos.xlsx](#)**.

CREACIÓN ONTOLOGÍA

Para llevar a cabo este paso, empleé la herramienta Protégé para la creación de las clases que consideré más pertinentes para la realización de la tarea propuesta. En total, se generaron 59 clases y subclases.

Las clases principales que utilicé incluyen:

- **Animal:** Representa la categoría principal, con subclases correspondientes a animales, cuyas URIs se extrajeron de la Ontología de Taxonomía de Vertebrados (<https://bioportal.bioontology.org/ontologies/VTO>).
- **Antibody:** Clase destinada a recoger las instancias de anticuerpos comerciales, cada uno con su URI propia creada para su identificación.
- **Application:** Representa técnicas de ensayos de bioquímica y biología molecular, con subclases que poseen URIs propias creadas para su clasificación.
- **Cell:** Engloba nombres de células, cuyas subclases se derivan de la Ontología Celular (<https://bioportal.bioontology.org/ontologies/CL>).
- **Clonality:** Define la clonabilidad de los anticuerpos como monoclonales o policlonales, organizando así las instancias según esta característica.
- **Company:** Representa las casas comerciales que fabrican los anticuerpos, con subclases dirigidas a URIs específicas para cada empresa.
- **Organelle:** Sus subclases identifican regiones celulares donde se localiza el antígeno detectado por cada anticuerpo. Las URIs fueron obtenidas de Componentes Celulares de la Gene Ontology (<https://bioportal.bioontology.org/ontologies/GO>).
- **Protein:** Clase con URI de Uniprot (<http://purl.uniprot.org/core/Protein>) destinada a contener instancias de proteínas reconocidas por los anticuerpos.
- **Storage:** Define el tipo de almacenamiento recomendado para cada anticuerpo.

Esta estructura ontológica proporciona un marco organizativo para la información recopilada sobre anticuerpos comerciales, permitiendo una fácil navegación y consulta por parte de los usuarios.

CREACIÓN DE LOS INDIVIDUOS

En el proceso inicial, se asignó una URI exclusiva a cada anticuerpo, seguido de la creación de otro individuo utilizando dicha URI, que albergaba toda la información específica de ese anticuerpo. Posteriormente, se estableció la propiedad "owl:sameAs" para vincular estas dos instancias.

Los pasos subsiguientes fueron los siguientes:

- 1- **Tipo de Anticuerpo:** Se dotó al individuo del tipo "Antibody" mediante la propiedad "rdf:type".
- 2- **Dilución Utilizada:** Se proporcionó información sobre la dilución utilizada para ese anticuerpo mediante la propiedad "rdfs:comment", especificando la dilución utilizada en mi contexto particular.
- 3- **Citas en la Bibliografía:** A través de la propiedad creada "citations", se registró el número de citas que tiene ese anticuerpo en la bibliografía, obteniendo esta información de la web (<https://www.citeab.com/>).
- 4- **Animal de Origen:** Se utilizó la propiedad "host" para indicar el animal en el que se creó el anticuerpo.
- 5- **Reactividad en Animales:** A través de la propiedad "reactivity_in", se señalaron los animales en los que se ha observado que el anticuerpo funciona, basándose en la bibliografía.
- 6- **Clonación:** La propiedad "has_clonality" se empleó para indicar si el anticuerpo es policlonal o monoclonal.
- 7- **Empresa Comercial:** Mediante la propiedad "has_company", se indicó la compañía comercial que fabrica el anticuerpo.
- 8- **Almacenamiento:** La propiedad "storage" se utilizó para especificar el lugar recomendado para almacenar el anticuerpo.
- 9- **Ensayo Aplicado:** La propiedad "applies_to" se empleó para indicar el ensayo en el que se ha probado el anticuerpo, basándose en la bibliografía.
- 10- **Instancia de Proteína:** Se creó (si no se había creado previamente) la instancia de la proteína, extrayendo la URI de Uniprot (<http://purl.uniprot.org/uniprot/>).
- 11- **Detección de Antígeno:** A través de la propiedad "detects_antigen", se estableció la conexión entre el anticuerpo y la proteína que reconoce.

12- **Localización Celular:** Utilizando la URI de la proteína, y mediante las propiedades "located_in" de Cell Ontology (<https://bioportal.bioontology.org/ontologies/CL>) y Gene Ontology (<https://bioportal.bioontology.org/ontologies/GO>) se indicaron las células y los compartimentos celulares donde se encuentra la proteína, respectivamente.

Una vez acabado las métricas del dataset fueron:

| | |
|----------------------------------|-----|
| <i>Axiom</i> | 676 |
| <i>Logical axiom count</i> | 385 |
| <i>Declaration axioms count</i> | 126 |
| <i>Class count</i> | 59 |
| <i>Object property count</i> | 11 |
| <i>Individual count</i> | 55 |
| <i>Annotation Property count</i> | 3 |

El documento para protegé lo puedes encontrar en **ESD/Ontology.rdf**.

VALIDACIÓN DEL DATASET

Una vez exportados el dataset en formato turtle (<https://www.w3.org/TR/turtle/>), procedí a organizar el archivo para escribir los prefix.

- @prefix owl: <<http://www.w3.org/2002/07/owl#>> .
- @prefix rdf: <<http://www.w3.org/1999/02/22-rdf-syntax-ns#>> .
- @prefix xml: <<http://www.w3.org/XML/1998/namespace>> .
- @prefix xsd: <<http://www.w3.org/2001/XMLSchema#>> .
- @prefix rdfs: <<http://www.w3.org/2000/01/rdf-schema#>> .
- @prefix obo: <<http://purl.obolibrary.org/obo/>> .
- @prefix mionto: <<https://dayhoff.inf.um.es/ontologies/anticuerpos#>> .
- @prefix metobo: <<https://data.bioontology.org/metadata/obo/>> .
- @prefix up: <<http://purl.uniprot.org/core/>> .
- @prefix uniprot: <<https://www.uniprot.org/uniprotkb/>> .

El document en formato turtle lo puedes encontrar en **ESD/AntibodyStorageTurtle**.

Una vez organizado el dataset en formato turtle, procedí a convertirlo a formato N-triples mediante la herramienta (<https://www.easyrdf.org/converter>). El documento en formato RDF/XML lo puedes encontrar en **ESD/AntibodyStorage rdfXml.txt**

Valido el texto en formato RDF/XML con la herramienta (<https://www.w3.org/RDF/Validator/>). Observamos que hay un total de 539 tripletas. La imagen del grafo lo puedes encontrar en **ESD/grafico.gif**

Vuelvo a convertir el documento a N-triples para luego generar los N-quads para meterlo a blazegraph. A este documento le añado al final de cada tripleta una URI del grafo. Este documento lo puedes encontrar en **ESD/AntibodyStorageNquad.nq**.

En blazegraph creo un namespace llamado AntibodyDataset. En la pestaña UPDATE, selecciono mi documento AntibodyStorage.nq y lo subo en formato N-quads. Una vez cargado el documento, selecciono los metadatos y los cargo en formato turtle. En los metadatos, en el namegraph, podrás ver el mismo nombre que se sitúa en el documento de AntibodyStorage.nq para generar el grafo. El documento de los metadatos lo podrás encontrar en **ESD/AntibodyStorage metadata**.

CONSULTAS SPARQL

Query 1:

Buscar todas las proteínas y todos los anticuerpos asociados a esas proteínas.

```
PREFIX rdfs:    <http://www.w3.org/2000/01/rdf-schema#>
PREFIX mionto:  <https://dayhoff.inf.um.es/ontologies/anticuerpos#>
PREFIX up:      <http://purl.uniprot.org/core/>
PREFIX uniprot: <https://www.uniprot.org/uniprotkb/>

SELECT ?protein ?proteinID ?antibody
WHERE {
    ?proteinID rdfs:type up:Protein.
    ?proteinID rdfs:label ?protein.
    ?o mionto:detects_antigen ?proteinID .
```

```

    ?o rdfs:label ?antibody
}

```

| protein | proteinID | antibody |
|-------------------|---|------------------------|
| Aqp4 protein | https://www.uniprot.org/uniprotkb/P55087 | Aqp4 Milipore |
| ASC protein | https://www.uniprot.org/uniprotkb/Q9ULZ3 | ASC SantaCruz |
| Brn3a Protein | https://www.uniprot.org/uniprotkb/Q01851 | Brn3a Milipore |
| Brn3a Protein | https://www.uniprot.org/uniprotkb/Q01851 | Brn3a SantaCruz |
| Brn3a Protein | https://www.uniprot.org/uniprotkb/Q01851 | Brn3a Synaptic Systems |
| Caspase 1 protein | https://www.uniprot.org/uniprotkb/P29466 | Caspase 1 Abcam |
| Caspase 3 protein | https://www.uniprot.org/uniprotkb/P42574 | Caspase 3 SantaCruz |
| CFH protein | https://www.uniprot.org/uniprotkb/P08603 | CFH ABclonal |
| Gfap protein | https://www.uniprot.org/uniprotkb/P14136 | Gfap Abcam |
| Gfap protein | https://www.uniprot.org/uniprotkb/P14136 | Gfap Milipore |
| Gfap protein | https://www.uniprot.org/uniprotkb/P14136 | Gfap Synaptic Systems |
| Iba1 protein | https://www.uniprot.org/uniprotkb/P55008 | Iba1 Abcam |
| Iba1 protein | https://www.uniprot.org/uniprotkb/P55008 | Iba1 Abcam 2 |
| MLKL protein | https://www.uniprot.org/uniprotkb/Q8NB16 | MLKL Abcam |
| Opn3 protein | https://www.uniprot.org/uniprotkb/Q9H1Y3 | Opsin Red/Green |
| P2X7 protein | https://www.uniprot.org/uniprotkb/Q99572 | P2X7 SantaCruz |
| Rbpms protein | https://www.uniprot.org/uniprotkb/Q93062 | Rbpms Genetex |
| Recoverin protein | https://www.uniprot.org/uniprotkb/P35243 | Recoverin Milipore |
| TMEM119 protein | https://www.uniprot.org/uniprotkb/Q8R138 | TMEM119 Abcam |
| TNFR1 protein | https://www.uniprot.org/uniprotkb/P19438 | TNFR1 Milipore |

Query 2:

Buscar los anticuerpos que detecten antígenos presentes en células gliales. En pantalla mostrar el nombre del anticuerpo, en qué célula se detecta, el host donde se ha hecho el anticuerpo y la URL a la web con toda la información del anticuerpo.

PREFIX owl: <http://www.w3.org/2002/07/owl#>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX obo: <http://purl.obolibrary.org/obo/>

PREFIX mionto: <https://dayhoff.inf.um.es/ontologies/anticuerpos#>

SELECT ?Antibodyname ?Cell ?Host ?AntibodyID

WHERE {


```
{?ab rdf:type mionto:Antibody.
?ab rdfs:label ?Antibodyname .
?ab owl:sameAs ?AntibodyID .
?ab mionto:host ?h .
?h rdfs:label ?Host .
?ab mionto:detects_antigen ?protein.
?protein obo:RO_0001025 obo:CL_0000129.
obo:CL_0000129 rdfs:label ?Cell}
```

UNION

```
{?ab rdf:type mionto:Antibody.
?ab rdfs:label ?Antibodyname .
?ab owl:sameAs ?AntibodyID .
?ab mionto:host ?h .
?h rdfs:label ?Host .
?ab mionto:detects_antigen ?protein.
?protein obo:RO_0001025 obo:CL_0000127.
obo:CL_0000127 rdfs:label ?Cell}
```

UNION

```
{?ab rdf:type mionto:Antibody.
?ab rdfs:label ?Antibodyname .
?ab owl:sameAs ?AntibodyID .
?ab mionto:host ?h .
?h rdfs:label ?Host .
?ab mionto:detects_antigen ?protein.
?protein obo:RO_0001025 obo:CL_0000636 .
obo:CL_0000636 rdfs:label ?Cell}
```

}

| Antibodyname | Cell | Host | AntibodyID |
|-----------------------|-------------|--------------------------------|---|
| Iba1 Abcam | Microglia | Capra (goat) | https://www.abcam.com/products/primary-antibodies/iba1-antibody-ab5076.html |
| Iba1 Abcam 2 | Microglia | Oryctolagus cuniculus (rabbit) | https://www.abcam.com/products/primary-antibodies/iba1-antibody-eor16588-ab178846.html |
| TMEM119 Abcam | Microglia | Oryctolagus cuniculus (rabbit) | https://www.abcam.com/products/primary-antibodies/tmem119-antibody-28-3-microglial-marker-ab209064.html |
| Gfap Abcam | Astrocyte | Capra (goat) | https://www.abcam.com/products/primary-antibodies/gfap-antibody-ab53554.html |
| Gfap Milipore | Astrocyte | Oryctolagus cuniculus (rabbit) | https://www.sigmaaldrich.com/DE/en/product/sigma/g9269 |
| Gfap Synaptic Systems | Astrocyte | Cavia porcellus (guinea pig) | https://www.sysv.com/product/173004 |
| Gfap Abcam | Müller cell | Capra (goat) | https://www.abcam.com/products/primary-antibodies/gfap-antibody-ab53554.html |
| Gfap Milipore | Müller cell | Oryctolagus cuniculus (rabbit) | https://www.sigmaaldrich.com/DE/en/product/sigma/g9269 |
| Gfap Synaptic Systems | Müller cell | Cavia porcellus (guinea pig) | https://www.sysv.com/product/173004 |

Query 3

Buscar un anticuerpo para utilizarlo en citometría de flujo para un ensayo para detectar el núcleo de las células de la microglía. Mostrar el nombre de la proteína, el host donde se ha hecho el anticuerpo, la clonalidad, la reactividad (en ratón), la dilución utilizada para el anticuerpo y dónde está guardado.

```
PREFIX rdfs:    <http://www.w3.org/2000/01/rdf-schema#>
PREFIX obo:     <http://purl.obolibrary.org/obo/>
PREFIX mionto:  <https://dayhoff.inf.um.es/ontologies/anticuerpos#>

SELECT ?protein ?host ?clonality ?reactivity ?dilution ?storage
WHERE {
    ?s rdfs:type mionto:Antibody .
    ?s mionto:applies_to mionto:FlowCitometry . # applies to Flow citometry
    ?s mionto:detects_antigen ?p .
    ?p rdfs:label ?protein .
    ?p obo:RO_0001025 obo:GO_0005634 . # located_in nucleus
    ?s mionto:host ?h .
    ?h rdfs:label ?host .
    ?s mionto:has_clonality ?clo .
    ?clo rdfs:label ?clonality .
    ?s mionto:reactivity_in obo:VTO_0014661 . # reactivity in mouse
    obo:VTO_0014661 rdfs:label ?reactivity .
    ?s rdfs:comment ?dilution .
    ?s mionto:storaged_in ?sto .
    ?sto rdfs:label ?storage
}
```

| protein | host | clonality | reactivity | dilution | storage |
|--------------|--------------------------------|------------|----------------------|----------------------|---------|
| Iba1 protein | Oryctolagus cuniculus (rabbit) | Polyclonal | Mus musculus (mouse) | usage at 1:500 in IF | freezer |

Query 4

Buscar el anticuerpo con el máximo de citas en la bibliografía de cada proteína. También extraer la media de citas en la bibliografía de los anticuerpos y el número de anticuerpos asociado a cada proteína.

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX mionto: <https://dayhoff.inf.um.es/ontologies/anticuerpos#>

```
SELECT ?protein (MAX(?c) AS ?MAXcitaciones) (xsd:integer(AVG(?c)) AS ?mediacitas)
(COUNT(?s) AS ?anticuerpos_proteina)
```

```
WHERE {
```

```
    ?s rdfs:type mionto:Antibody .
```

```
    ?s mionto:citations ?c .
```

```
    ?s mionto:detects_antigen ?prot .
```

```
    ?prot rdfs:label ?protein .
```

```
}
```

```
GROUP BY ?protein
```

```
HAVING (xsd:integer(COUNT(?s)) >= 2)
```

```
ORDER BY DESC(?MAXcitaciones)
```

| protein | MAXcitaciones | mediacitas | anticuerpos_proteina |
|---------------|---------------|------------|----------------------|
| Iba1 protein | 1493 | 1016 | 2 |
| Gfap protein | 459 | 329 | 3 |
| Brn3a Protein | 267 | 163 | 3 |

Query 5

Hacer una consulta federada para buscar información de la proteína MLKL de nuestro dataset en la Uniprot. Extraer la información del nombre, mnemonico, si está revisada la proteína, la última modificación en la base de datos de esa proteína, las secuencias que tenga la proteína y el número de participaciones en GO.

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX up: <http://purl.uniprot.org/core/>

```
SELECT DISTINCT ?protein ?nombre ?mnemonic ?review ?fechamodificacion ?seq
(COUNT(?GO) AS ?numeroGO)
```

```
WHERE {
```

```
    ?protein rdfs:type up:Protein .
```

```
    ?protein rdfs:label "MLKL protein"
```

```
    SERVICE <https://sparql.uniprot.org/> {
```

```
        ?protein rdfs:label ?nombre .
```

```
        ?protein up:mnemonic ?mnemonic.
```

```
        ?protein up:reviewed ?review .
```

```
        ?protein up:modified ?fechamodificacion .
```

```
        ?protein up:sequence ?seq .
```

```
        ?protein up:classifiedWith ?i .
```

```
        ?i rdfs:label ?GO
```

```
    }
```

```
}
```

```
GROUP BY ?protein ?nombre ?mnemonic ?review ?fechamodificacion ?seq
```

| protein | nombre | mnemonic | review | fechamodificacion | seq | numeroGO |
|---|--|------------|--------|-------------------|---|----------|
| http://purl.uniprot.org/uniprot/Q8NB16 | Mixed lineage kinase domain-like protein | MLKL_HUMAN | true | 2023-11-08 | http://purl.uniprot.org/isoforms/Q8NB16-2 | 96 |
| http://purl.uniprot.org/uniprot/Q8NB16 | Mixed lineage kinase domain-like protein | MLKL_HUMAN | true | 2023-11-08 | http://purl.uniprot.org/isoforms/Q8NB16-1 | 96 |

BÚSQUEDAS DESDE EL PAQUETE SPARQL DE R

Estas mismas consultas que realicé desde blazegraph las utilizo para hacer las consultas desde R con el paquete SPARQL. A continuación dispongo el código para hacer las consultas.

library(SPARQL)

```
endpoint <- " http://34.69.252.26:3031/blazegraph/namespace/AntibodyDataset/sparql"
```

#Query 1

```
query1 <- "PREFIX rdfs:  <http://www.w3.org/2000/01/rdf-schema#>
```

```
PREFIX mionto: <https://dayhoff.inf.um.es/ontologies/anticuerpos#>
```

```
PREFIX up:      <http://purl.uniprot.org/core/>
```

```
PREFIX uniprot: <https://www.uniprot.org/uniprotkb/>
```

```
SELECT ?protein ?proteinID ?antibody
```

```
WHERE {
```

```
    ?proteinID rdf:type up:Protein.
```

```
    ?proteinID rdfs:label ?protein.
```

```
    ?o mionto:detects_antigen ?proteinID .
```

```
    ?o rdfs:label ?antibody
```

```
}"
```

```
resq1<- SPARQL(endpoint,query1)
```

```
resq1$results
```

#Query 2

```
query2 <- "PREFIX owl:  <http://www.w3.org/2002/07/owl#>
```

```
PREFIX rdfs:  <http://www.w3.org/2000/01/rdf-schema#>
```

```
PREFIX obo:    <http://purl.obolibrary.org/obo/>
```

```
PREFIX mionto: <https://dayhoff.inf.um.es/ontologies/anticuerpos#>
```

```
SELECT ?Antibodyname ?Cell ?Host ?AntibodyID
```

```
WHERE {
```

```
    {?ab rdf:type mionto:Antibody.
```

```
    ?ab rdfs:label ?Antibodyname .
```

```
    ?ab owl:sameAs ?AntibodyID .
```

```
    ?ab mionto:host ?h .
```

```
    ?h rdfs:label ?Host .
```

```
    ?ab mionto:detects_antigen ?protein.
```

```
?protein obo:RO_0001025 obo:CL_0000129.
```

```
obo:CL_0000129 rdfs:label ?Cell}
```

```
UNION
```

```
{?ab rdf:type mionto:Antibody.
```

```
?ab rdfs:label ?Antibodyname .
```

```
?ab owl:sameAs ?AntibodyID .
```

```
?ab mionto:host ?h .
```

```
?h rdfs:label ?Host .
```

```
?ab mionto:detects_antigen ?protein.
```

```
?protein obo:RO_0001025 obo:CL_0000127.
```

```
obo:CL_0000127 rdfs:label ?Cell}
```

```
UNION
```

```
{?ab rdf:type mionto:Antibody.
```

```
?ab rdfs:label ?Antibodyname .
```

```
?ab owl:sameAs ?AntibodyID .
```

```
?ab mionto:host ?h .
```

```
?h rdfs:label ?Host .
```

```
?ab mionto:detects_antigen ?protein.
```

```
?protein obo:RO_0001025 obo:CL_0000636 .
```

```
obo:CL_0000636 rdfs:label ?Cell}
```

```
}"
```

```
resq2<- SPARQL(endpoint,query2)
```

```
resq2$results
```

```
#Query 3
```

```
query3 <- "PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
```

```
PREFIX obo: <http://purl.obolibrary.org/obo/>
```

```
PREFIX mionto: <https://dayhoff.inf.um.es/ontologies/anticuerpos#>
```

```
SELECT ?protein ?host ?clonality ?reactivity ?dilution ?storage
```

```
WHERE {
```

```
    ?s rdf:type mionto:Antibody .
```

```
?s mionto:applies_to mionto:FlowCitometry . # applies to Flow cytometry
```

```

?s mionto:detects_antigen ?p .
?p rdfs:label ?protein .

?p obo:RO_0001025 obo:GO_0005634 . # located_in nucleus

?s mionto:host ?h .
?h rdfs:label ?host .

?s mionto:has_clonality ?clo .
?clo rdfs:label ?clonality .

?s mionto:reactivity_in obo:VTO_0014661 . # reactivity in mouse
obo:VTO_0014661 rdfs:label ?reactivity .

?s rdfs:comment ?dilution .

?s mionto:stored_in ?sto .
?sto rdfs:label ?storage
}"

```

```
resq3<- SPARQL(endpoint,query3)
```

```
resq3$results
```

#Query 4

```

query4 <- "PREFIX rdfs:  <http://www.w3.org/2000/01/rdf-schema#>
PREFIX mionto:  <https://dayhoff.inf.um.es/ontologies/anticuerpos#>

```

```

SELECT ?protein (MAX(?c) AS ?MAXcitaciones) (xsd:integer(AVG(?c)) AS ?mediacitas) (COUNT(?s) AS
?anticuerpos_proteina)

```

```
WHERE {
```

```

?s rdf:type mionto:Antibody .

```

```

?s mionto:citations ?c .

```

```

?s mionto:detects_antigen ?prot .

```

```

?prot rdfs:label ?protein .

```

```

}
```

```
GROUP BY ?protein
```

```
HAVING (xsd:integer(COUNT(?s)) >= 2)
```

```
ORDER BY DESC(?MAXcitaciones)"
```

```
resq4<- SPARQL(endpoint,query4)
```

```
resq4$results
```

```
#Query 5
```

```
query5 <- "PREFIX rdfs:  <http://www.w3.org/2000/01/rdf-schema#>
```

```
PREFIX up:                <http://purl.uniprot.org/core/>
```

```
SELECT DISTINCT ?protein ?nombre ?mnemonic ?review ?fechamodificacion ?seq (COUNT(?GO) AS  
?numeroGO)
```

```
WHERE {
```

```
?protein rdf:type up:Protein .
```

```
?protein rdfs:label 'MLKL protein'
```

```
SERVICE <https://sparql.uniprot.org/> {
```

```
    ?protein rdfs:label ?nombre .
```

```
    ?protein up:mnemonic ?mnemonic.
```

```
        ?protein up:reviewed ?review .
```

```
        ?protein up:modified ?fechamodificacion .
```

```
        ?protein up:sequence ?seq .
```

```
        ?protein up:classifiedWith ?i .
```

```
        ?i rdfs:label ?GO
```

```
    }
```

```
}
```

```
GROUP BY ?protein ?nombre ?mnemonic ?review ?fechamodificacion ?seq"
```

```
resq5<- SPARQL(endpoint,query5)
```

```
resq5$results
```

PUBLICACIÓN DEL DATASET

Para seguir con los principios FAIR (Findable, Accesible, Interoperable y Reusable) procedí a publicar los datos en pubby. Como ya hemos creado en blazegraph el

namespace con los datos y los metadatos, procedo a configurar los documentos para dirigir mi dataset para que pueda ser buscado por otros usuarios.

```
# Prefix declarations to be used in RDF output
@prefix conf: <http://richard.cyganiak.de/2007/pubby/config.rdf#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .

# Server configuration section
<> a conf:Configuration;
  # Project name for display in page titles
  conf:projectName "FAIR data publication UM MSc Bioinformatics";
  # Homepage with description of the project for the link in the page header
  conf:projectHomepage <https://github.com/mikel-egana-aranguren/UM-Bioinformatics-MSc-FAIR-data>;
  # The Pubby root, where the webapp is running inside the servlet container.
  conf:webBase <http://34.69.252.26:8181/>;
  # URL of an RDF file whose prefix mapping is to be used by the
  # server; defaults to <>, which is *this* file.
  conf:usePrefixesFrom <>;
  # If labels and descriptions are available in multiple languages,
  # prefer this one.
  conf:defaultLanguage "en";
  # When the homepage of the server is accessed, this resource will
  # be shown.
  conf:indexResource <https://github.com/mikel-egana-aranguren/UM-Bioinformatics-MSc-FAIR-data>;

# Dataset configuration
conf:dataset [
  # SPARQL endpoint URL of the dataset
  conf:sparqlEndpoint <http://34.69.252.26:3031/blazegraph/namespace/AntibodyDataset/sparql>;
  # Common URI prefix of all resource URIs in the SPARQL dataset
  conf:datasetBase <https://um.es/data/AntibodyDataset/>;
  # Will be appended to the conf:webBase to form the public
  # resource URIs; if not present, defaults to ""
  #conf:webResourcePrefix "resource/";
];
```

Anticuerpos comerciales para uso experimental

at FAIR data publication UM MSc Bioinformatics

<http://34.69.252.26:8181/>

| Property | Value |
|----------------|--|
| ?:License | ▪ http://creativecommons.org/licenses/MIT/ |
| ?:distribution | ▪ https://um.es/sparql/AntibodyDataset ▪ https://um.es/txt/ontology |
| ?:label | ▪ Anticuerpos comerciales para uso experimental () |
| ?:namedGraph | ▪ http://um.es/graph/AntibodyDataset |
| ?:primaryTopic | ▪ Anticuerpos () |
| ?:type | ▪ http://www.w3.org/ns/dcat#Dataset |

This page shows information obtained from the SPARQL endpoint at <http://34.69.252.26:3031/blazegraph/namespace/AntibodyDataset/sparql>.

[As Turtle](#) | [As RDF/XML](#) | [Browse in Disco](#) | [Browse in Tabulator](#) | [Browse in OpenLink Browser](#)