RDF KNOWLEDGE GRAPHS

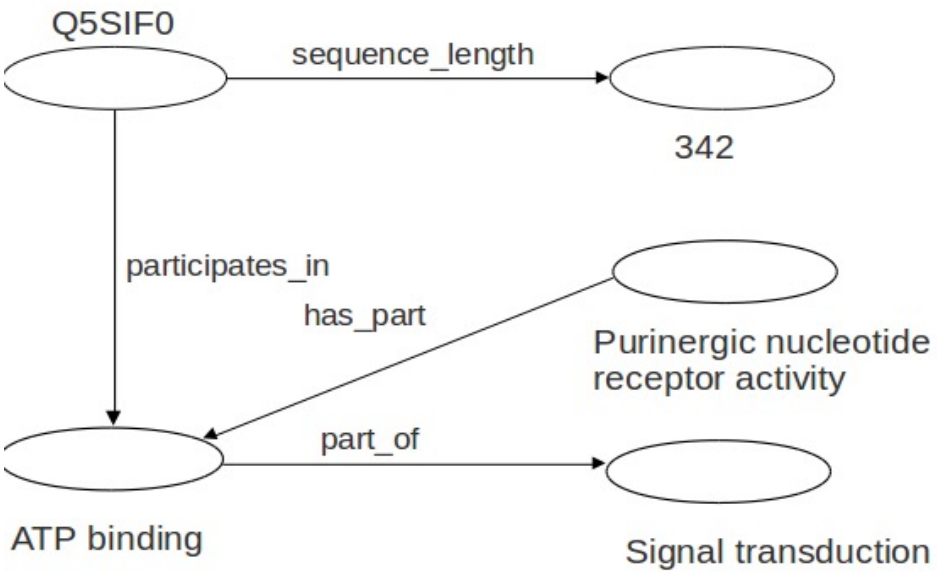
UNIVERSIDAD DE
MURCIA

RDF knowledge graphs

UNIVERSIDAD DE MURCIA

- Knowledge graphs

  DATA + KNOWLEDGE



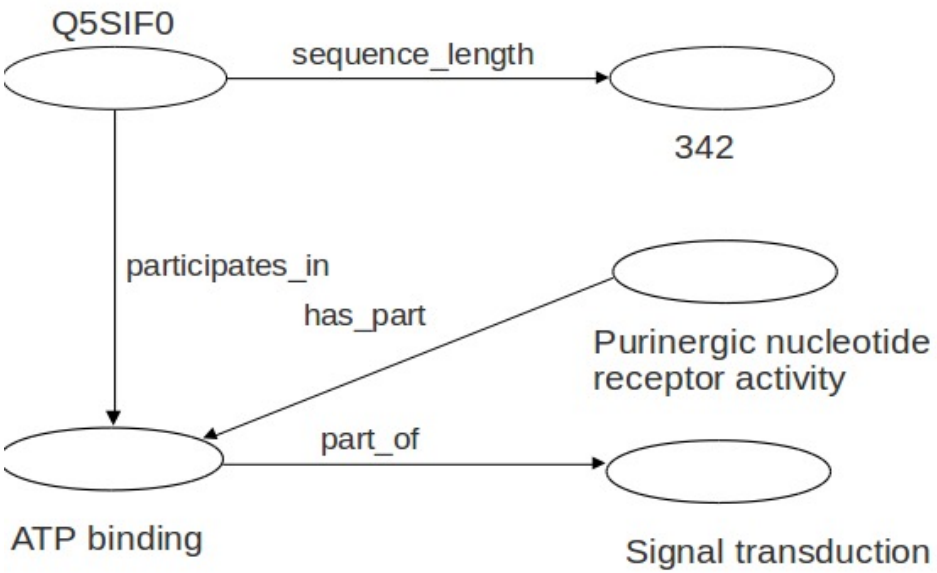
RDF knowledge graphs

UNIVERSIDAD DE MURCIA

- Knowledge graphs

  DATA + KNOWLEDGE

- How to represent the data? RDF

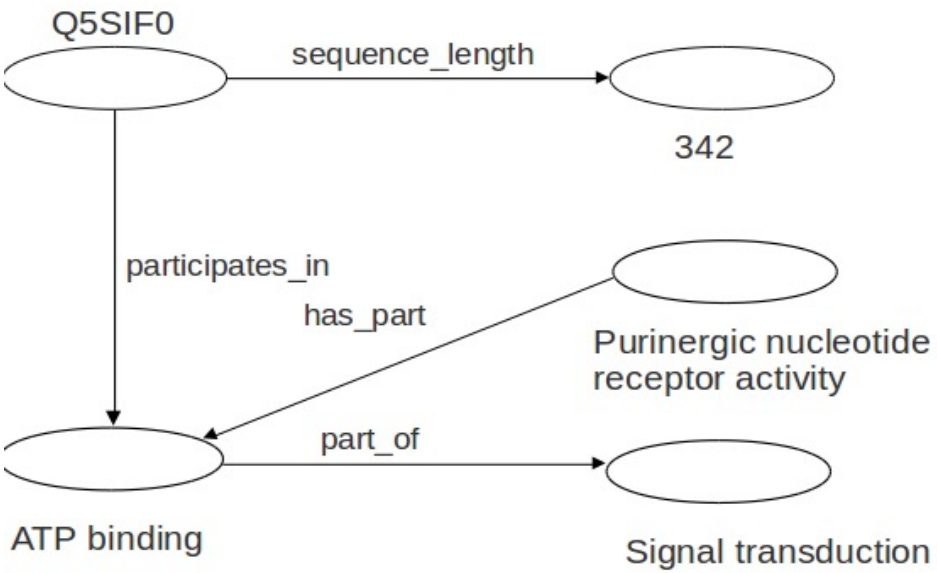


RDF knowledge graphs

UNIVERSIDAD DE MURCIA

- Knowledge graphs

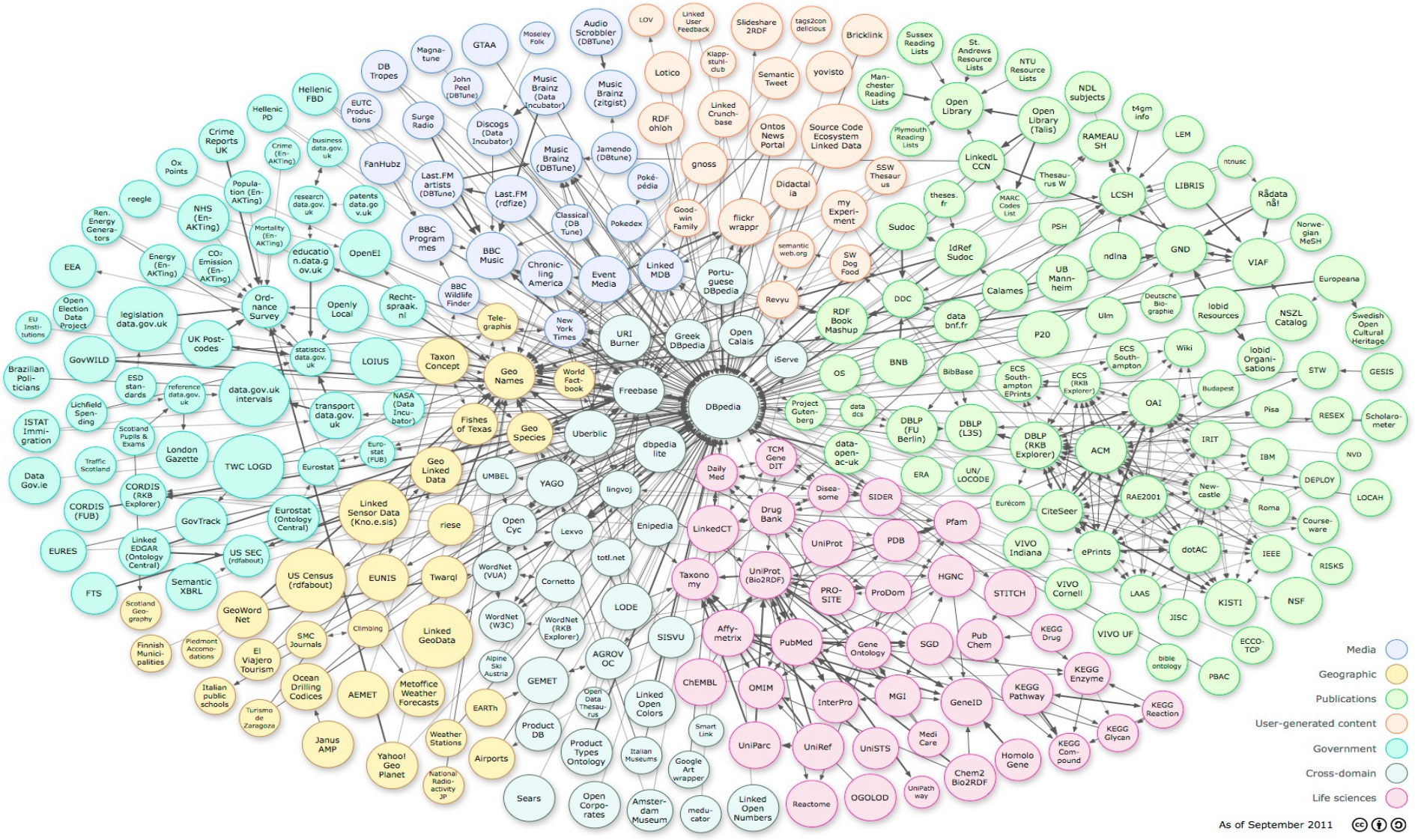
  DATA + KNOWLEDGE



- How to represent the data? RDF
- How to represent the knowledge? Ontologies

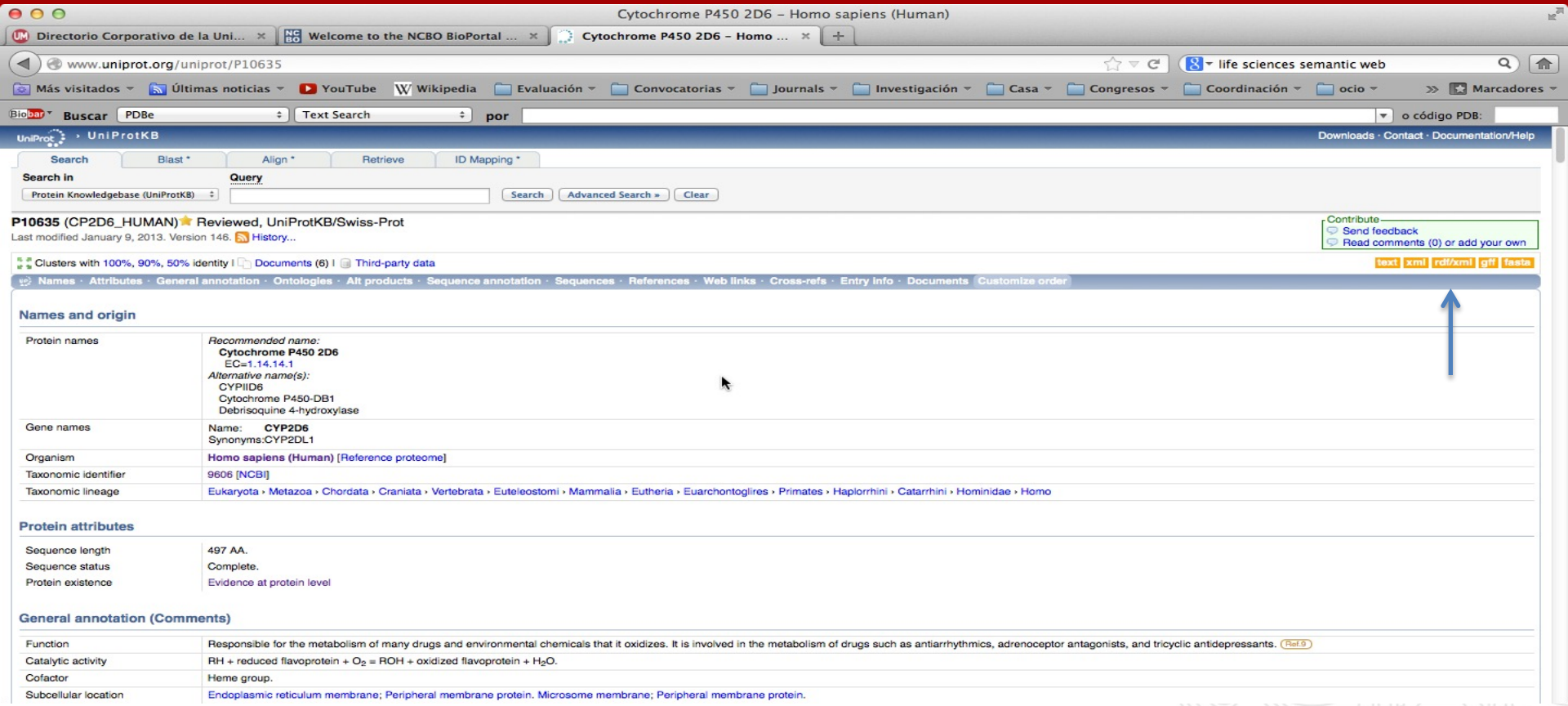
The Web of Data is at this level: RDF

UNIVERSIDAD DE MURCIA



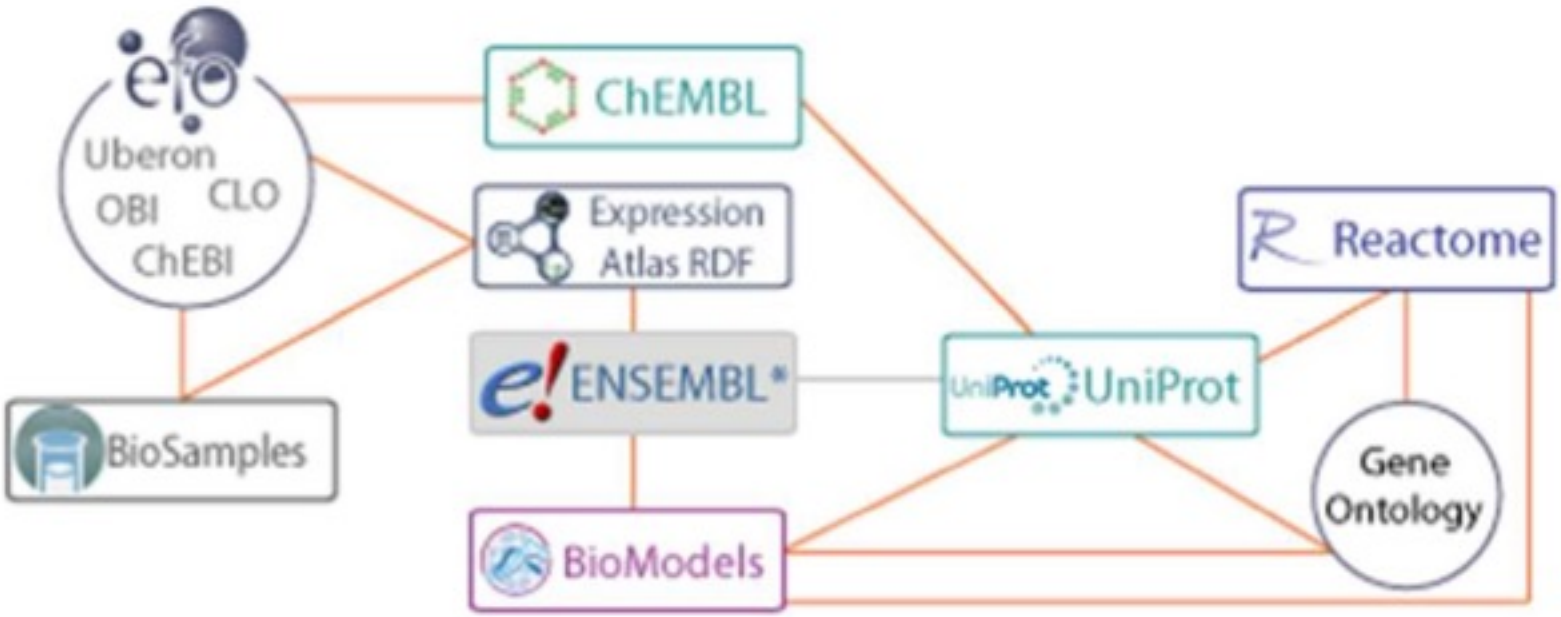
RDF in UniProt

UNIVERSIDAD DE MURCIA



EBI RDF Platform

UNIVERSIDAD DE MURCIA



http://www.ebi.ac.uk/rdf/

Swiss Bioinformatics Institute

UNIVERSIDAD DE MURCIA

SIB RDF Data

UNIVERSIDAD DE MURCIA

Table 1.
SIB databases providing free, linked open data for reuse

Database	SPARQL endpoint URL	Type of data
Bgee	https://www.bgee.org/sparql/ https://purl.org/bioquery (Bio-Query)	Gene expression
Cellosaurus	https://query.wikidata.org/ (via Wikidata)	Cell line
GlyConnect	https://beta.glyconnect.expasy.org/sparqlsweetshttps://glyconnect.expasy.org/sparql (only machine-readable)	Glycoprotein
HAMAP	https://hamap.expasy.org/sparql	Protein family classification and annotation rules
MetaNetX	https://rdf.metanetx.org/	Metabolic network
OMA	https://sparql.omabrowser.org/	Orthologous protein-coding gene
OrthoDB	https://sparql.orthodb.org/	Orthologous protein-coding gene
Rhea	https://sparql.rhea-db.org/	Enzymatic and transport reaction
STRING	https://sparql.string-db.org/	Protein-protein interactions
SwissLipids	https://beta.sparql.swisslipids.org/	Lipid
UniProtKB	https://sparql.uniprot.org/	Protein

Bio2RDF
https://bio2rdf.org/

UNIVERSIDAD DE MURCIA

Linked Data for the Life Sciences

Select an ID example | Submit

about datasets download API SPARQL query repository mailing list

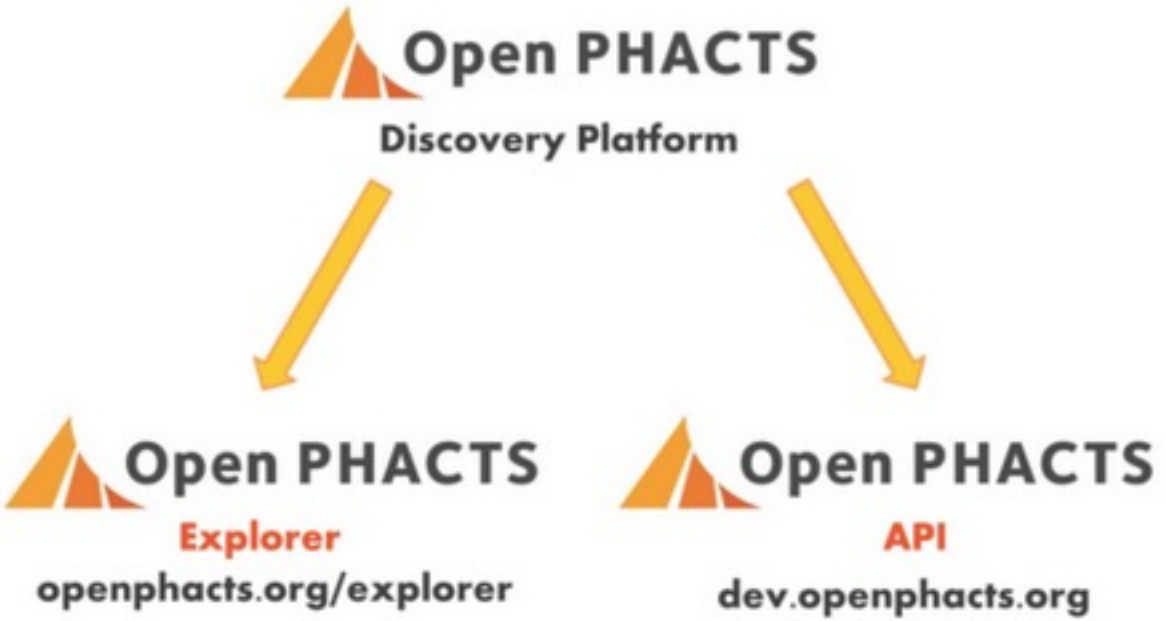
Powered by Docker, PHP, and Virtuoso Open-Source Edition

Maastricht University | Université Laval

Example: OpenPhacts
http://www.openphacts.org/

UNIVERSIDAD DE MURCIA

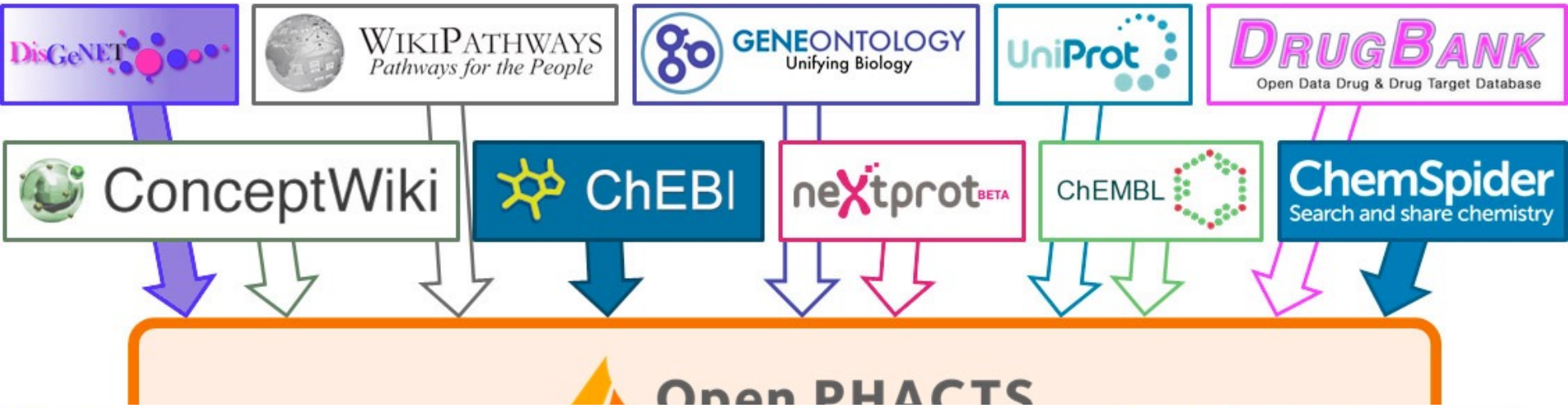
The Open PHACTS Discovery Platform has been developed to reduce barriers to drug discovery in industry, academia and for small businesses. You can now access the integrated pharmacological data via the Open PHACTS API, Open PHACTS Explorer or apps.



"Open PHACTS is enabling you to do what you do best!"

Data integration in OpenPhacts
http://www.openphacts.org/

UNIVERSIDAD DE MURCIA



Dataset	Downloaded	Version	Dataset Licence
ChEBI	04 Mar 2015	Release 125	CC-BY-SA 3.0
ChEMBL	18 Feb 2015	ChEMBL 20	CC-BY-SA 3.0
ConceptWiki	12 Dec 2013		
DisGeNET	31 Mar 2015	Version 2.1.0	Open Database License
DrugBank	19 Feb 2015	Version 4.1	Non-commercial use
ENZYME	02 Feb 2015	Release 2015_1	No modifications
FDA Adverse Events (FAERS)	09 July 2012		Public Domain (CC0)
Gene Ontology	04 Mar 2015		CC-BY 4.0
Gene Ontology Annotations	17 Feb 2015		CC-BY 4.0
neXTProt	01 Feb 2014	v1.0	CC-BY-ND 3.0
UniProt (manually curated entries)	28 Jan 2015	Release 2015_1	CC-BY-ND 3.0
WikiPathways	20 Mar 2015	v20150312	CC-BY 3.0

Discovery Apps in OpenPhacts
http://www.openphacts.org/

UNIVERSIDAD DE MURCIA

ChemBioNavigator

ChemBioNavigator allows you to query Open PHACTS and visualise groups of related molecules in chemical and biological space. For an introduction, take a look at BioSolveIT's recent webinar.

Developed by the University of Hamburg and BioSolveIT GmbH

ChemBioNavigator »

collector

Collector allows users to extract compounds annotated with experimental data from Open PHACTS, and build QSAR predictive models. Users can filter by target, biological data ranges and physicochemical properties.

Developed by PSMAR as part of the eTOX project

Collector »

COMBINE

COMBINE (COnstruct cheMical and BIological Network) can be used to build and visualise interactive chemical and biological networks. Users can create networks linking structures to targets to proteins to diseases to pathways.

Developed by CimplSoft

COMBINE »

GARField

GARField is an abbreviation for Graph-Activity-Relationship visualisation Field, and is a "polypharmacology browser" for exploring, visualising and predicting interactions between targets and compounds.

Developed by the Technical University of Denmark

GARField »

SciBites

SciBites provides a real-time drug discovery/pharma information portal, connecting the latest news on competitive intelligence for pharma and biotech companies directly to Open PHACTS pharmacology data.

Developed by SciBite Limited

SciBites »

Target Dossier

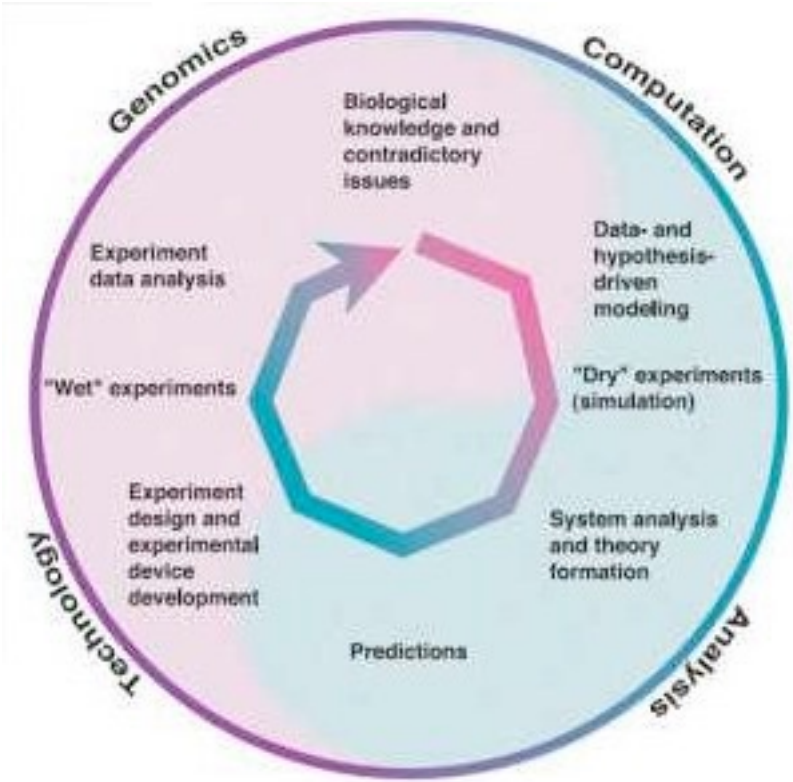
Target Dossier integrates data about drug targets to identify the most productive points for therapeutic intervention. Its user interfaces make use of Open PHACTS data to investigate and visualise target-centric services.

Developed by CNIO

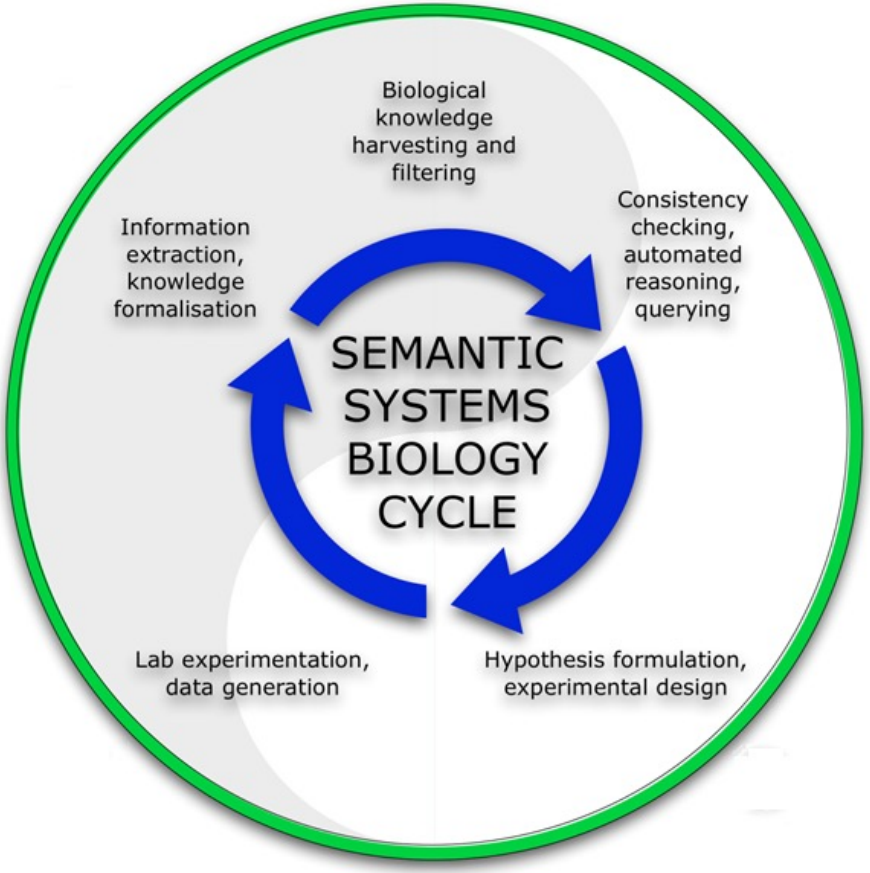
Target Dossier »

BioGateway (https://biogateway.eu)

UNIVERSIDAD DE MURCIA



Kitano, H. (2002). Computational systems biology. Nature, 420(6912), 206-210.

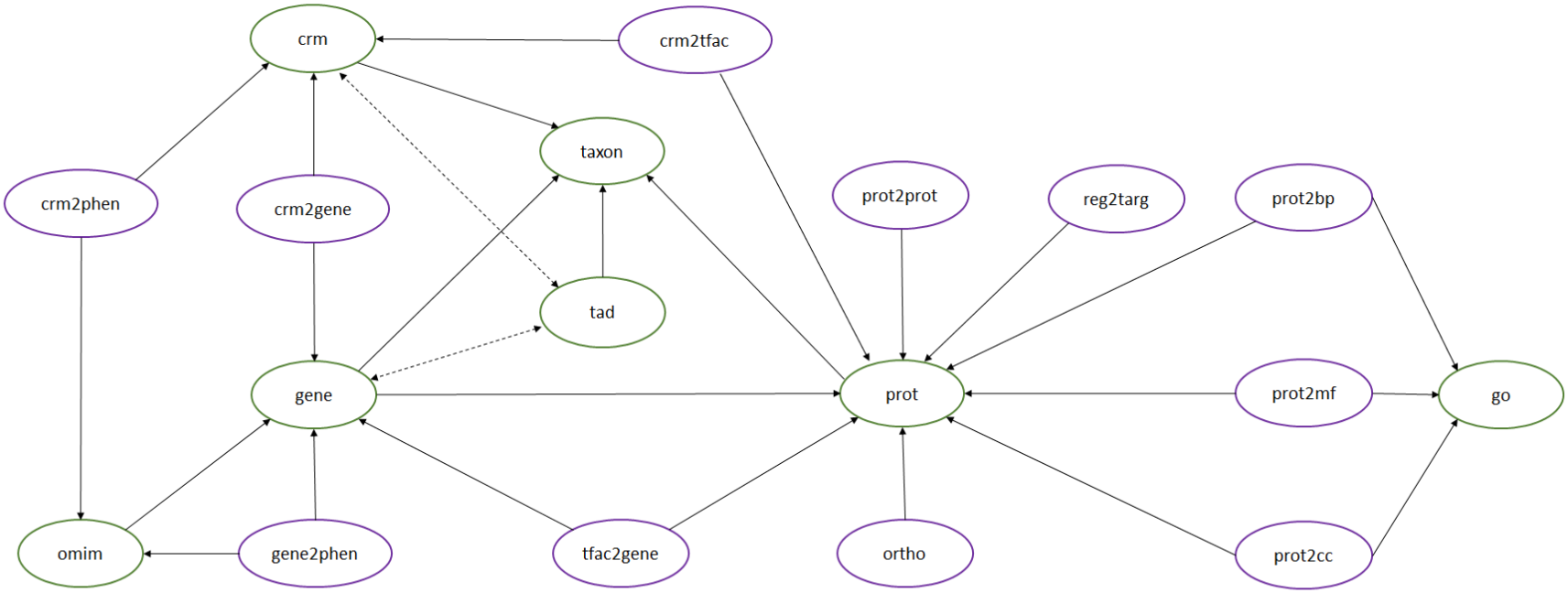


BioGateway

BioGateway (https://biogateway.eu)

UNIVERSIDAD DE MURCIA

hBGW network



crm: cis regulatory domains (enhancers)
tad: topologically associating domains
gene: protein-coding genes
prot: proteins
omim: Online Mendelian Inheritance in Man
go: Gene Ontology
taxon: NCBITaxon Ontology

links through properties

links through coordinate operations

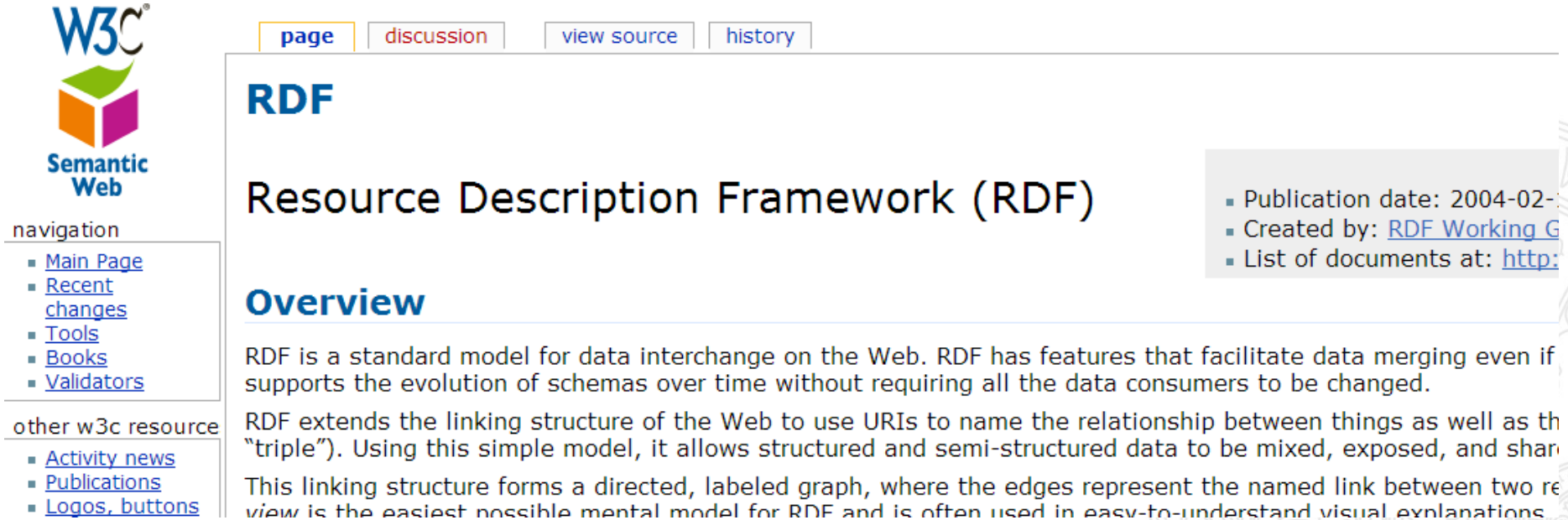
links through relation graphs

gene2phen: gene-phenotype relations
prot2prot: protein interactions
prot2bp: protein-biological process relations
prot2mf: protein-molecular function relations
prot2cc: protein-cellular component relations
tfac2gene: transcription factor-target gene relations
reg2targ: protein-protein regulatory relations
ortho: protein-protein orthology relations

RDF

UNIVERSIDAD DE MURCIA

https://www.w3.org/RDF/



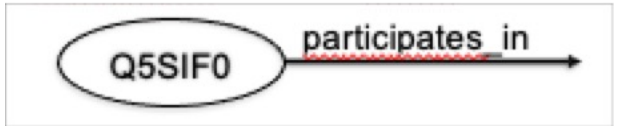
- RDF is graphical formalism (+ XML syntax + semantics)
  - for representing metadata
  - for describing the meaning of information in a machine-accessible way

RDF Data Model

UNIVERSIDAD DE MURCIA

- Three components:
  - **Resources:** any entity that can be described in RDF
    - Examples: web page, protein, gene, etc.

- **Properties:** characteristics, attributes or relations describing a resource. It can be a primitive value or a relation with another resource.



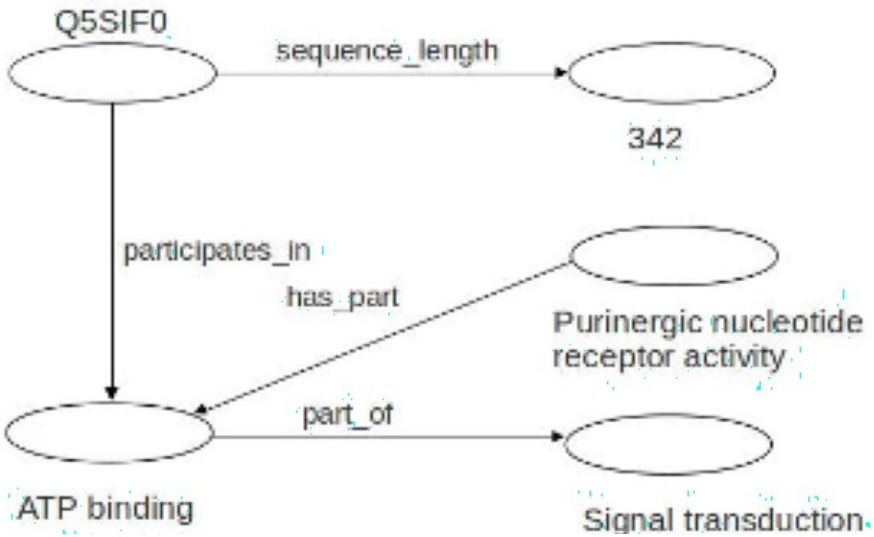
- **Statements or RDF triples (subject, predicate, object):** combination of:
  - The resource described **(Subject)**
  - The property or relation between the subject and the object **(Predicate)**
  - Value of the property **(Object),** which is a resource or a literal

RDF Data Model

- Example

  "The protein Q5SIF0 participates in the molecular function ATP binding

- RDF triples are combined forming graphs



A formal model for RDF

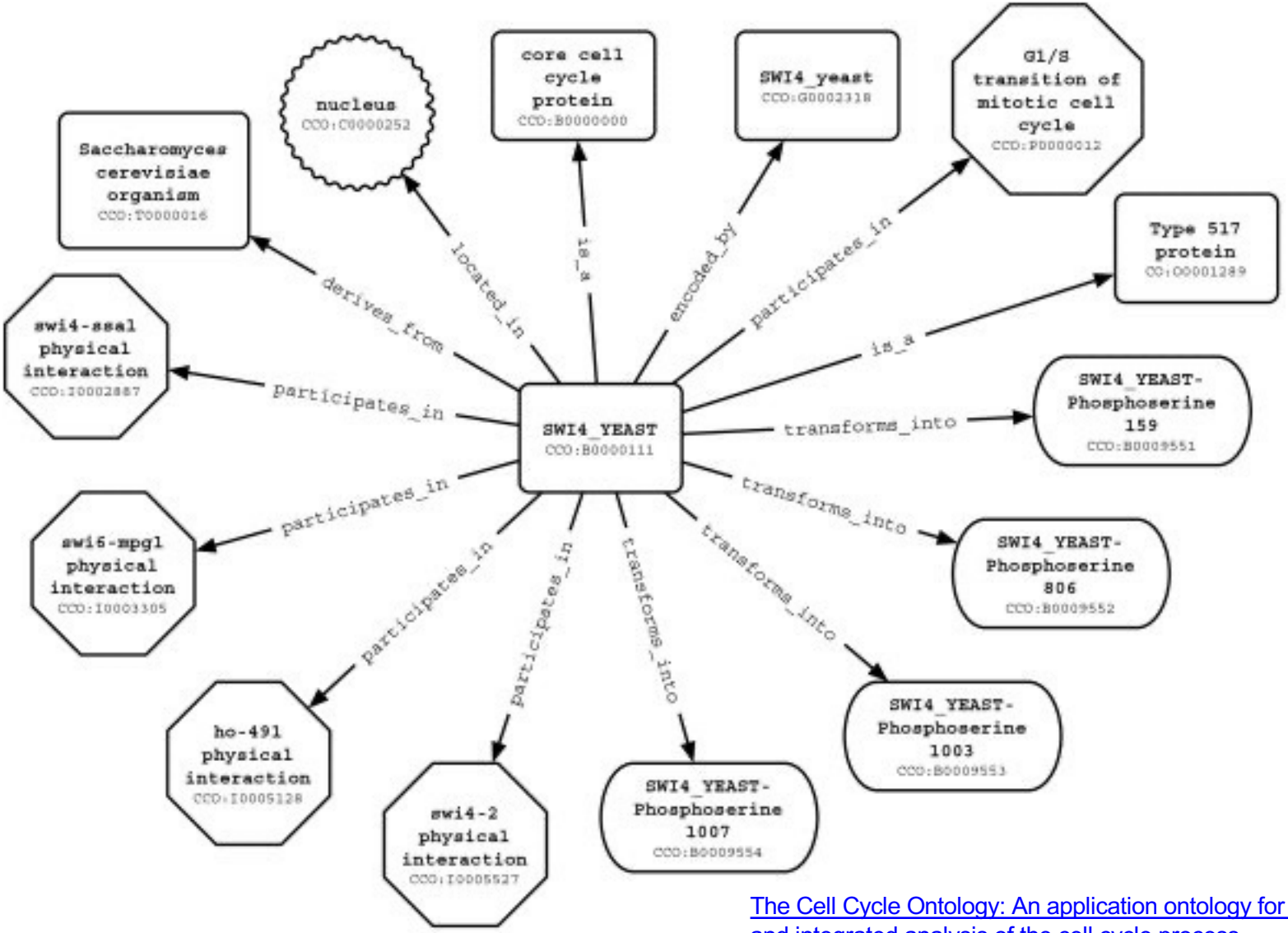
UNIVERSIDAD DE MURCIA

RDF itself is mathematically straightforward:

- Basic Definitions
  - **Resources.**
  - **Properties $\subset$ Resources called**
  - **Literals**
  - **Statements = Resources $\times$ Properties $\times$ {Resources $\cup$ Literals}**
- Typing
  - **rdf:type $\in$ Properties**
  - **{RDF:type, sub, obj} $\in$ Statements $\Rightarrow$ obj $\in$ Resources**

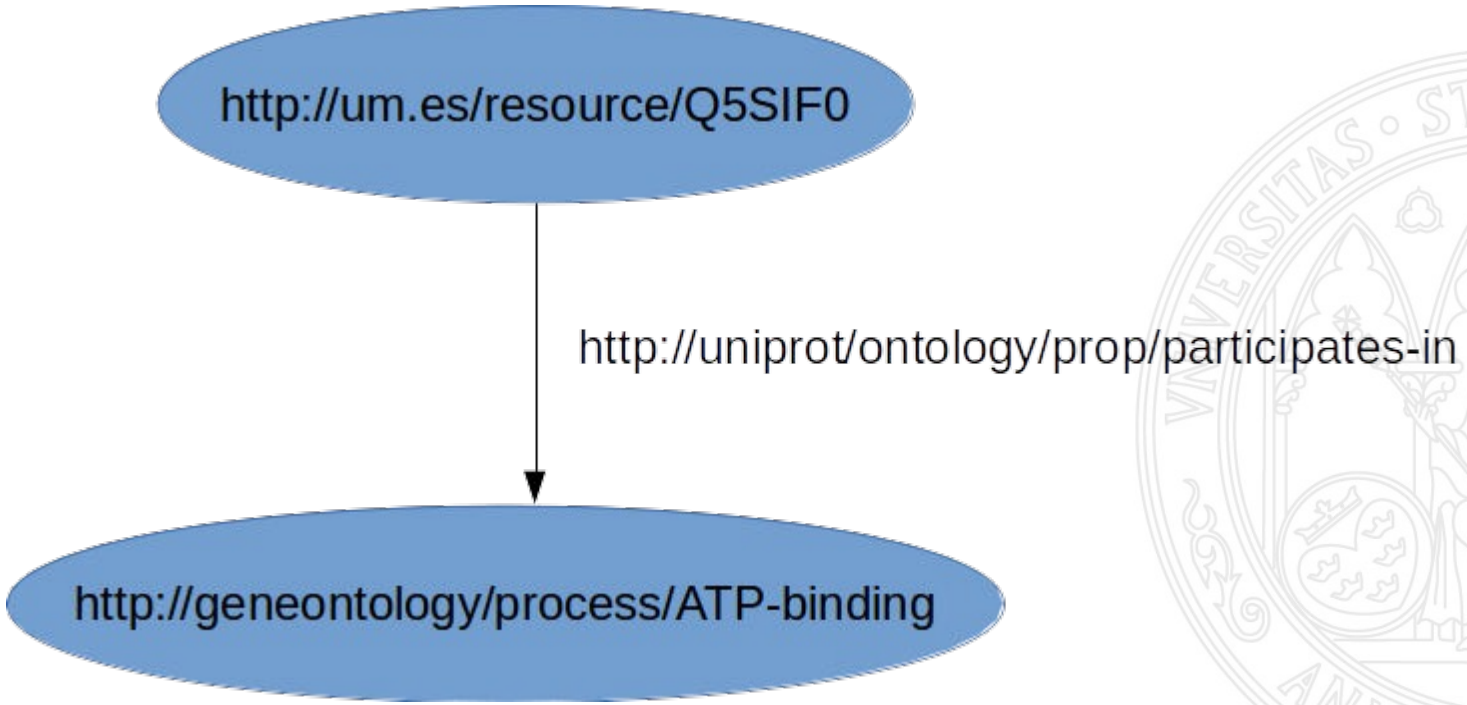
Example: the protein SWI4_YEAST

UNIVERSIDAD DE MURCIA



The Cell Cycle Ontology: An application ontology for the representation and integrated analysis of the cell cycle process

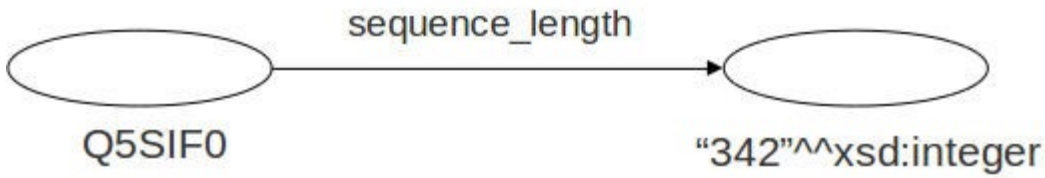
RDF Graph: URIs



RDF Graphs: content

UNIVERSIDAD DE MURCIA

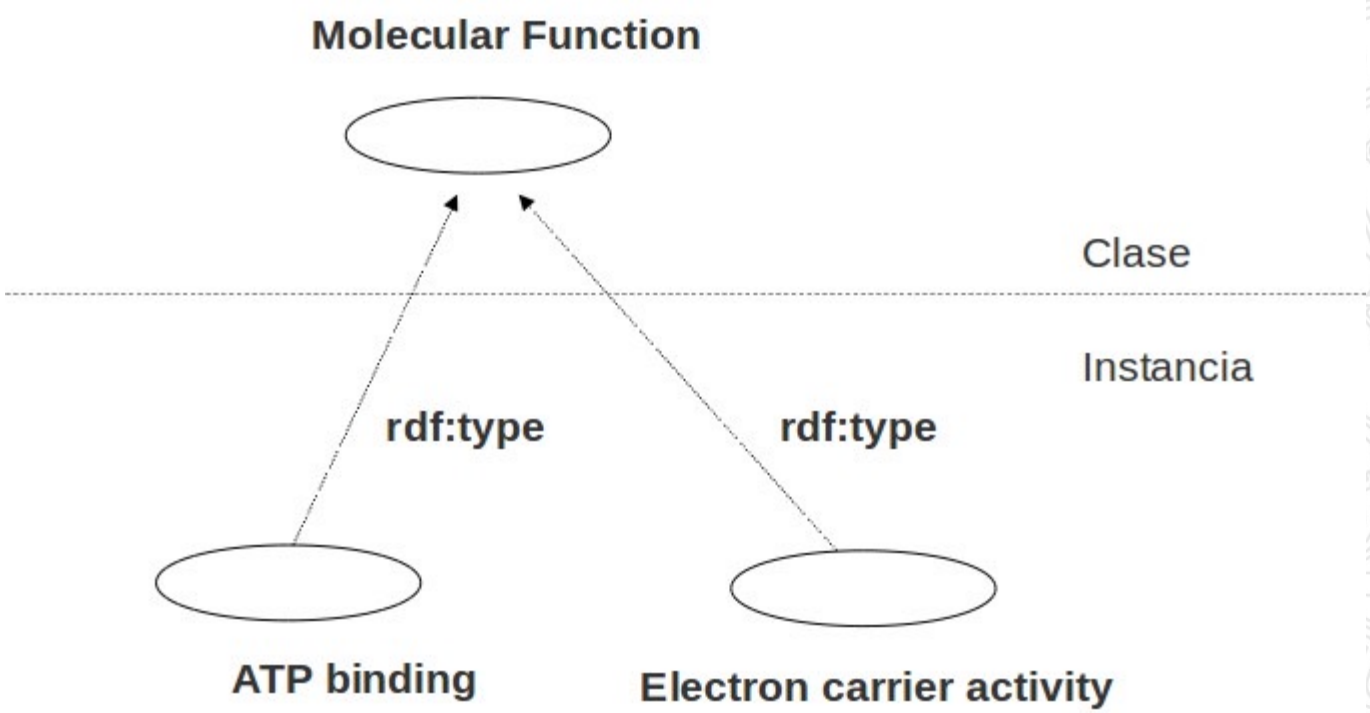
- The subjects and predicates can only be Resources
- Objects can be literal values
- The literal values may have a type (XML Schema Datatypes)



rdf:type

UNIVERSIDAD DE MURCIA

It allows grouping resources in classes, but there are no classes in RDF



RDF namespaces

UNIVERSIDAD DE MURCIA

- RDF uses namespaces to group URIs
- Namespaces can be abbreviated / expanded through PREFIX

PREFIX dbpedia: <http://dbpedia.org/resource>

dbpedia:Murcia = http://dbpedia.org/resource/murcia

…

Vocabularies

UNIVERSIDAD DE MURCIA

- Vocabulary: informally defined collection of URIs, usually in the same namespace

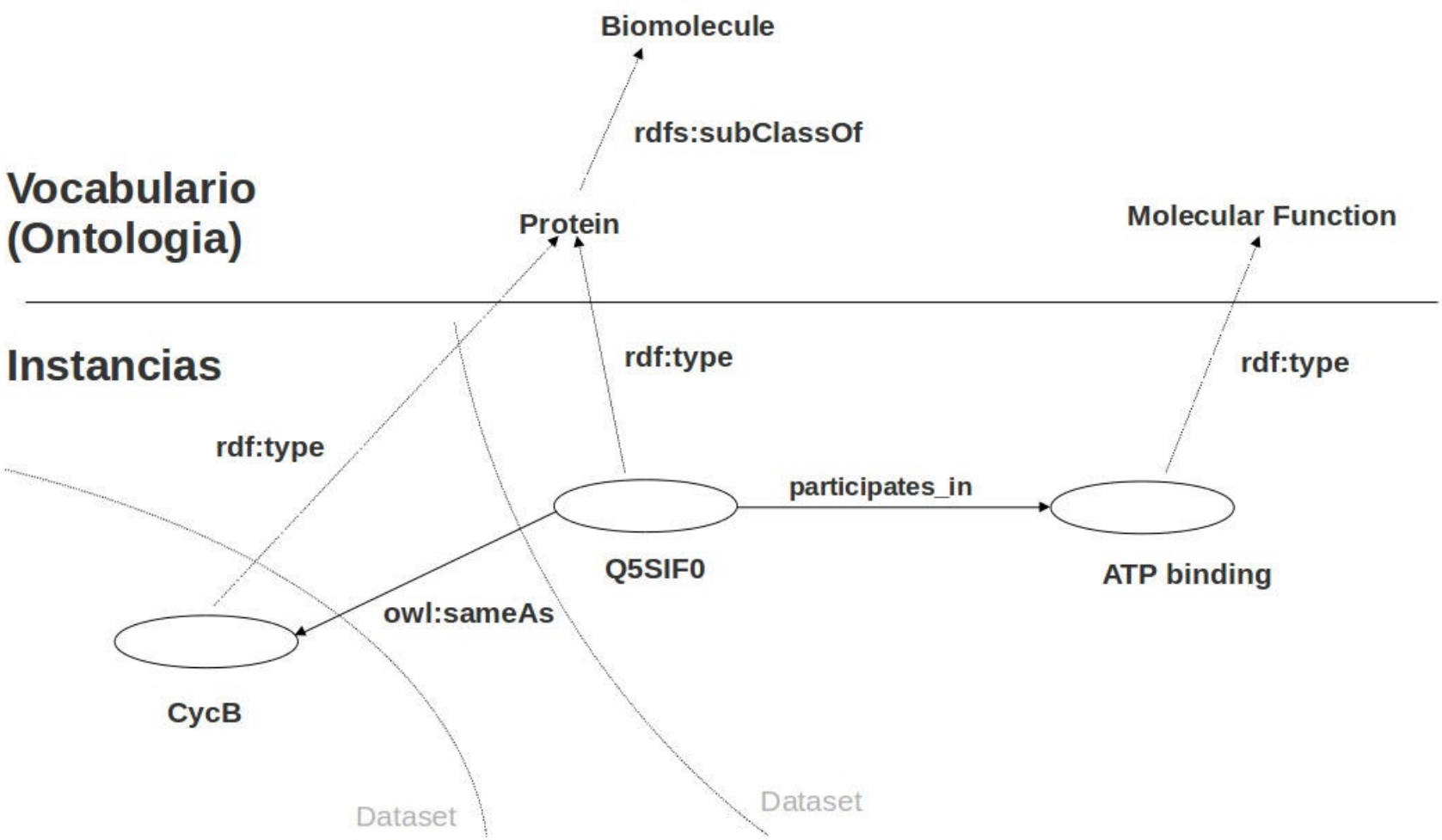
- There are “reserved” vocabularies (define languages)

RDF	http://www.w3.org/1999/02/22-rdf-syntax-ns#
RDFS	http://www.w3.org/2000/01/rdf-schema#
OWL	http://www.w3.org/2002/07/owl#

- Most vocabularies are ontologies that describe the general properties of the data we want to share and publish

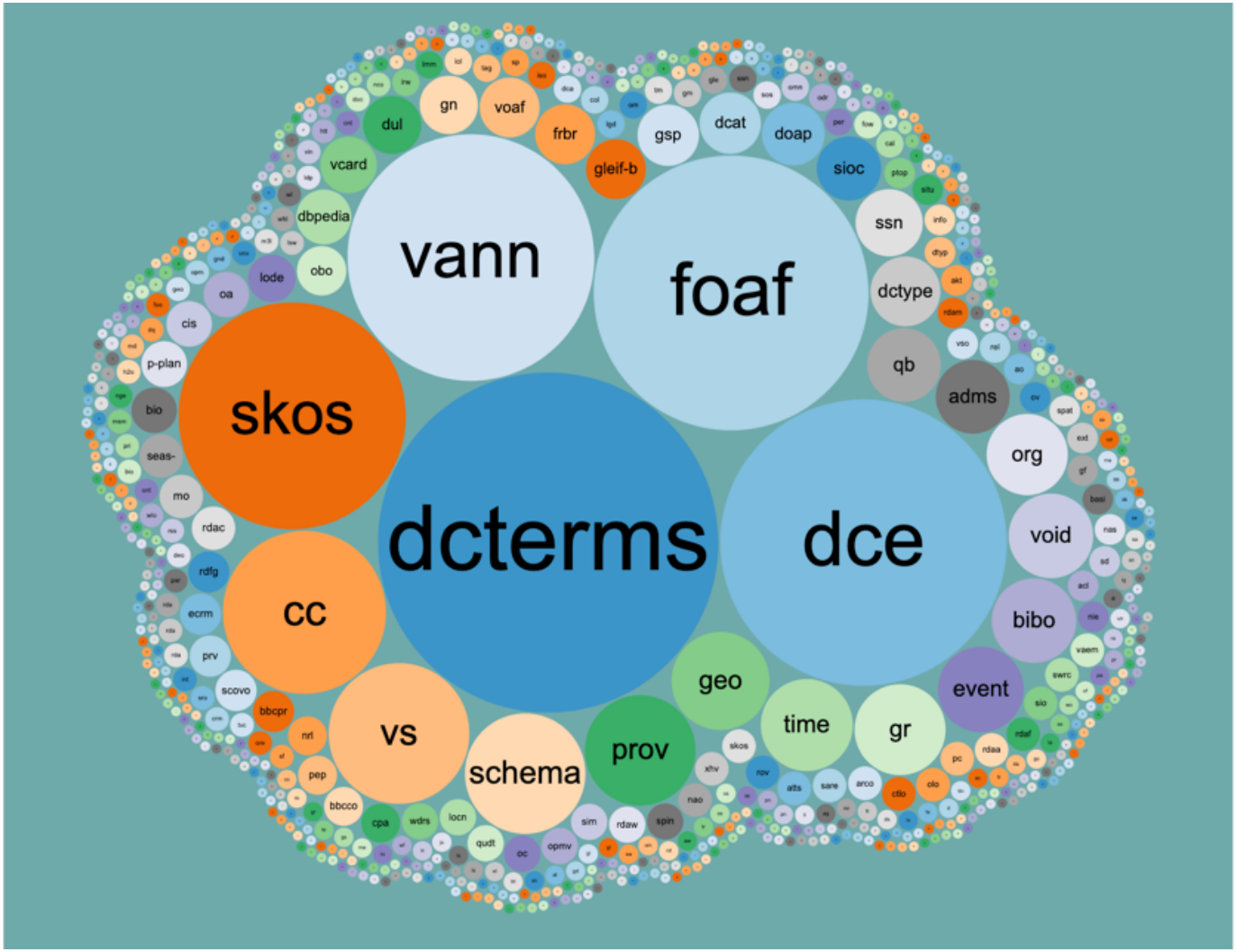
foaf:person
dbpedia-ont:city
dc:book ...

Ontologies


Biomolecule
rdfs:subClassOf
Vocabulario
(Ontologia)
Protein
Molecular Function
Instancias
rdf:type
rdf:type
rdf:type
participates_in
Q5SIF0
ATP binding
owl:sameAs
CycB
Dataset
Dataset

Linked Open Vocabularies
https://lov.linkeddata.es/dataset/lov/

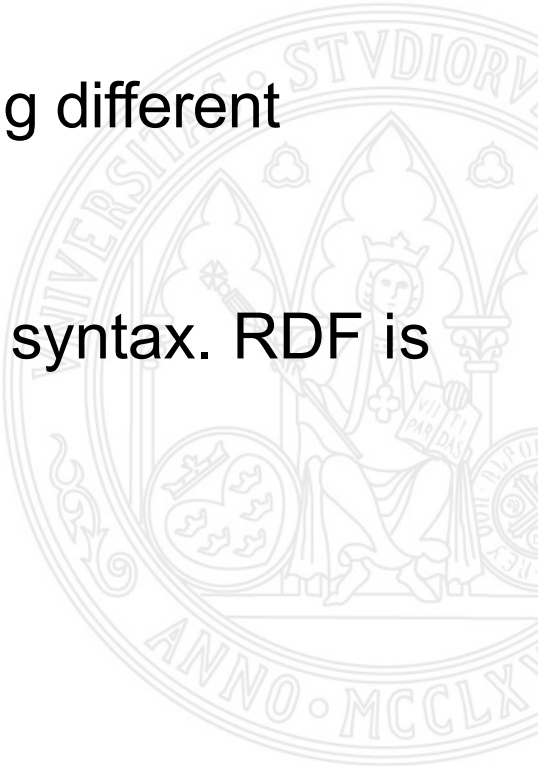
UNIVERSIDAD DE MURCIA



Data model ≠ syntax

UNIVERSIDAD DE MURCIA

- RDF is a data model
- The RDF data model can be expressed using different syntaxes, such as RDF/XML
- Please do NOT confuse the model with the syntax. RDF is much more than an XML file



RDF syntaxes

UNIVERSIDAD DE MURCIA

- Some of the most popular syntaxes:
  - RDF/XML (http://www.w3.org/TR/rdf-syntax-grammar/)
  - JSON-LD (https://www.w3.org/TR/json-ld/)
  - RDFa (http://www.w3.org/TR/rdfa-core/)
  - Turtle (http://www.w3.org/TR/turtle/)
- Useful tools:
  - http://www.w3.org/RDF/Validator/
  - http://www.easyrdf.org/converter

RDF syntaxes

UNIVERSIDAD DE MURCIA

- Pattern for RDF/XML documents. An RDF document contains:
  - One or more “descriptions” of resources
  - A description is a set of assertions about a resource
  - The element <*rdf:Description*> contains an atributte *rdf:about* which identifies the resource described; all the values must be URI.
  - All the properties of the resource are described in the child elements.

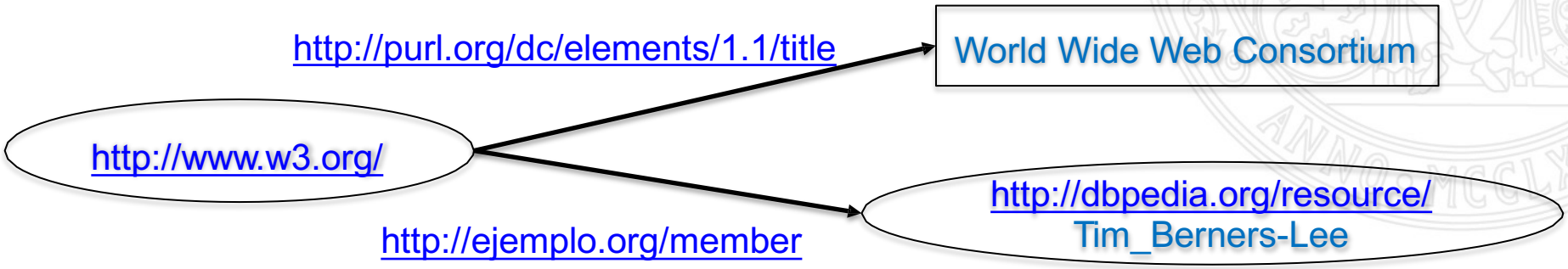
```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
      xmlns:X=”URI" >
<rdf:Description rdf:about=”URI">
      <X:Property 1>
      <X:Property 2>
</rdf:Description>
</rdf:RDF>
```

Example: RDF/XML Syntax

UNIVERSIDAD DE MURCIA

- http://www.w3.org/TR/rdf-syntax-grammar/

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:e="http://ejemplo.org/">
  <rdf:Description rdf:about="http://www.w3.org/">
    <dc:title>World Wide Web Consortium</dc:title>
    <e:member rdf:resource="http://dbpedia.org/resource/Tim_Berners-Lee"/>
  </rdf:Description>
</rdf:RDF>
```



Example: JSON-LD

UNIVERSIDAD DE MURCIA

- https://www.w3.org/TR/json-ld/

```
[
  {
    "@id":http://dbpedia.org/resource/Tim_Berners-Lee
  },
  {
    "@id":"http://www.w3.org/",
      http://purl.org/dc/elements/1.1/title:[
        {
          "@value":"World Wide Web Consortium”
        }
      ],
      "http://ejemplo.org/member":[
        {
          "@id":http://dbpedia.org/resource/Tim_Berners-Lee
        }
      ]
  }
]
```

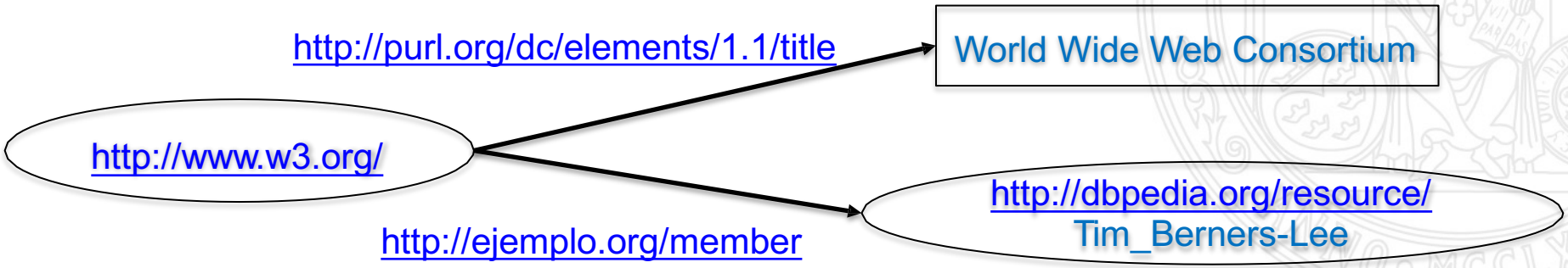
Example: Turtle

UNIVERSIDAD DE MURCIA

- http://www.w3.org/TR/turtle/

```
@prefix dc11: <http://purl.org/dc/elements/1.1/> .
@prefix ns0: <http://ejemplo.org/> .

<http://www.w3.org/>
      dc11:title "World Wide Web Consortium" ;
      ns0:member <http://dbpedia.org/resource/Tim_Berners-Lee> .
```



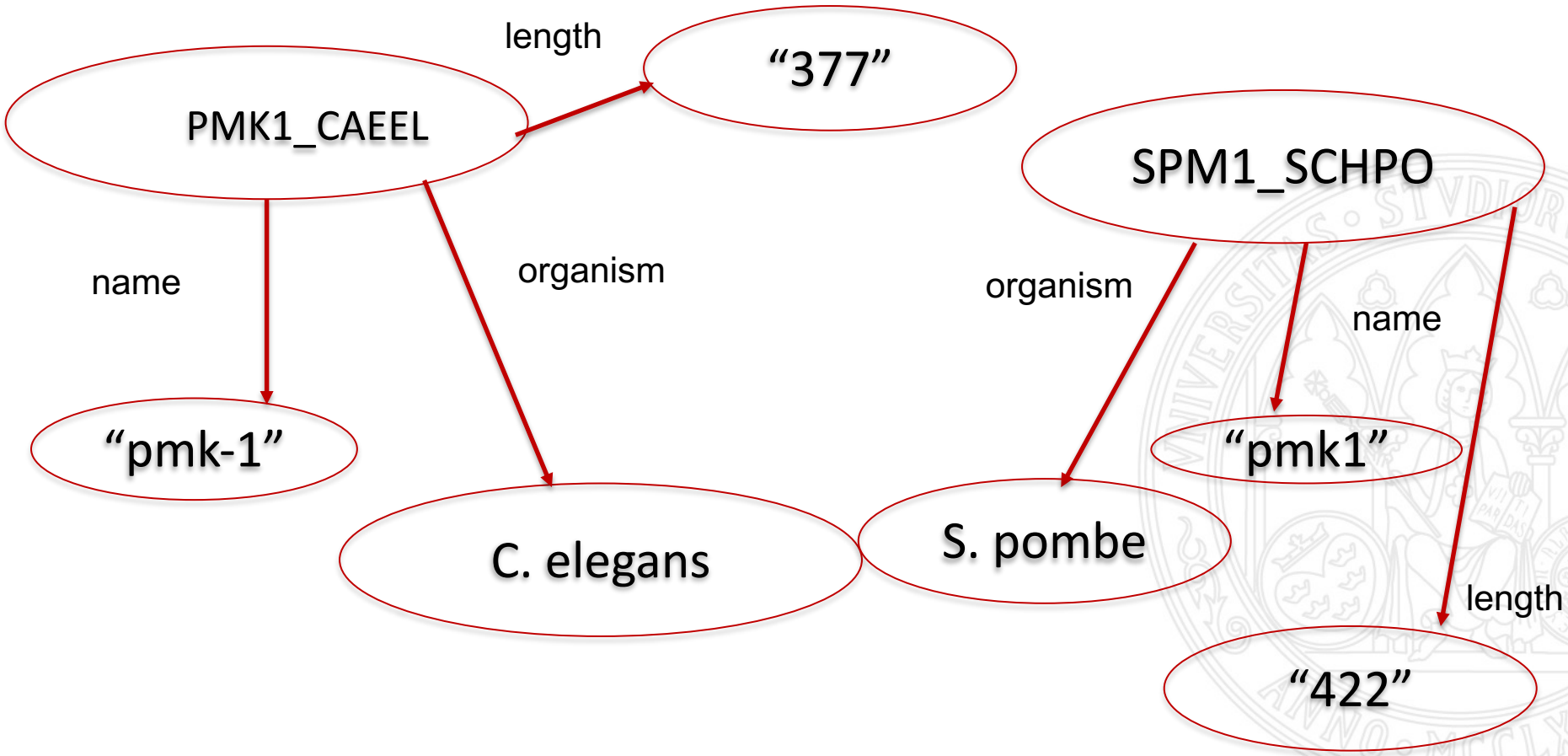
Exercise 1: Modeling proteins in RDF

UNIVERSIDAD DE MURCIA

Protein	Organism	Length
pmk-1	C. elegans	377
pmk1	S. pombe	422

Exercise 1: Modeling proteins in RDF

UNIVERSIDAD DE MURCIA



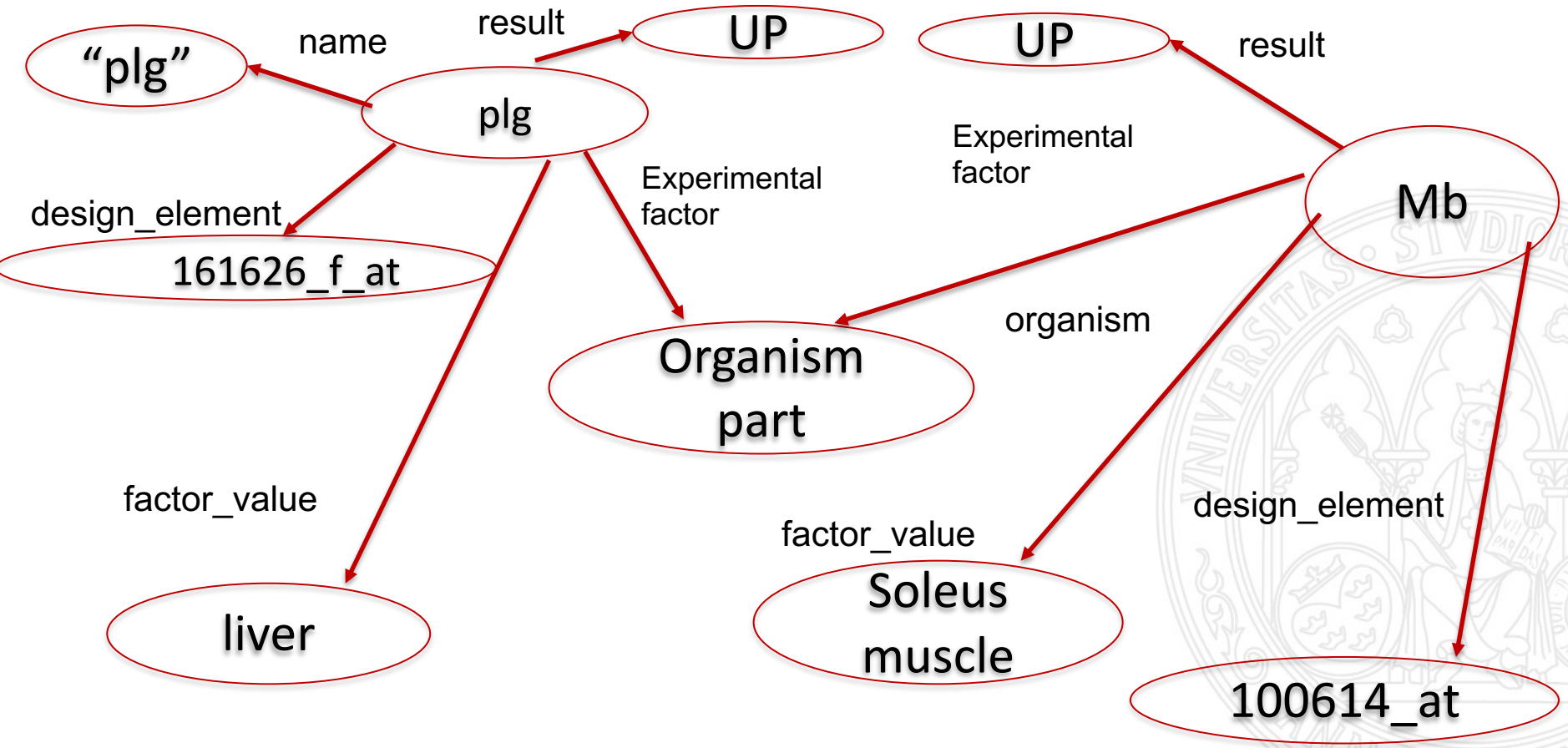
Exercise 2: Modeling gene expression experiments in RDF

UNIVERSIDAD DE MURCIA

Gene	Design elements	Experimental factor	Factor value	UP/DOWN
Plg	161626_f_at	Organism part	liver	UP
Mb	100614_at	Organism part	Soleus muscle	UP
Apoc2	97887_at	Organism part	liver	UP

Exercise 2: Modeling gene expression experiments in RDF

UNIVERSIDAD DE MURCIA



Reified statements

UNIVERSIDAD DE MURCIA

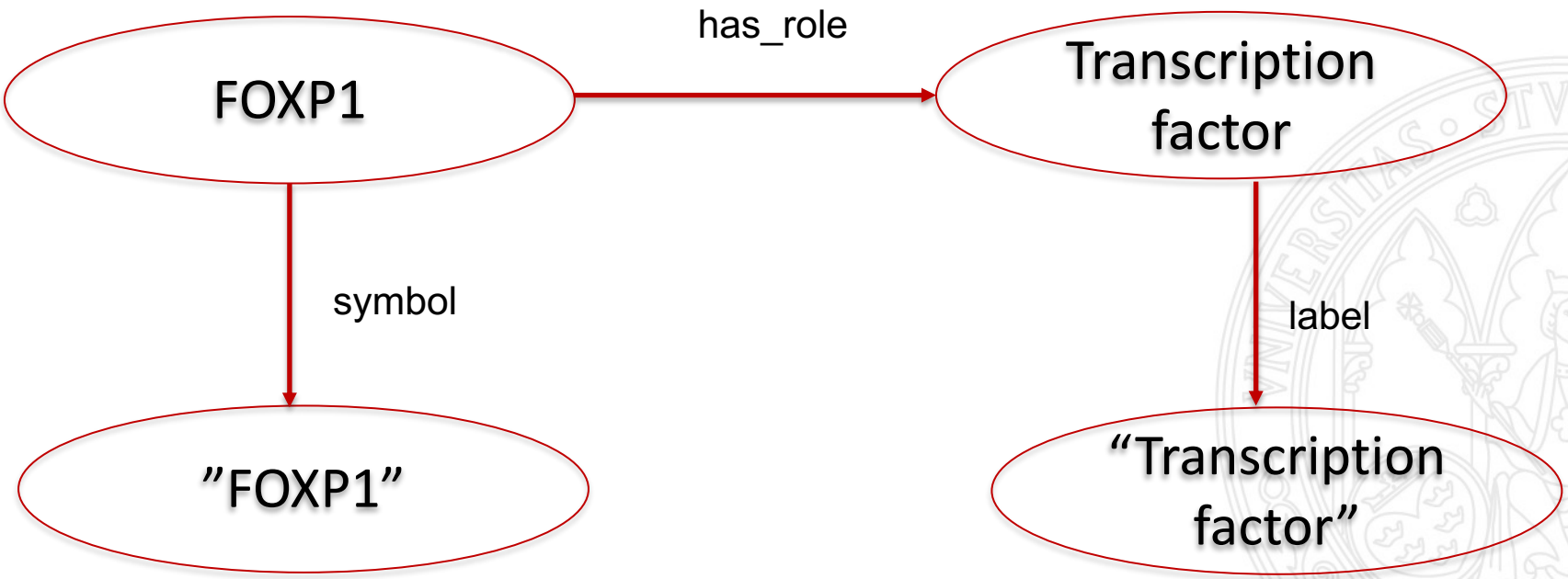
- **We can talk about things rather than asserting things**

- "FOXP1 plays the role of transcription factor" asserts a property of FOXP1

- "Jack believes that FOXP1 plays the role of transcription factor e" asserts a property of Jack, not of FOXP1

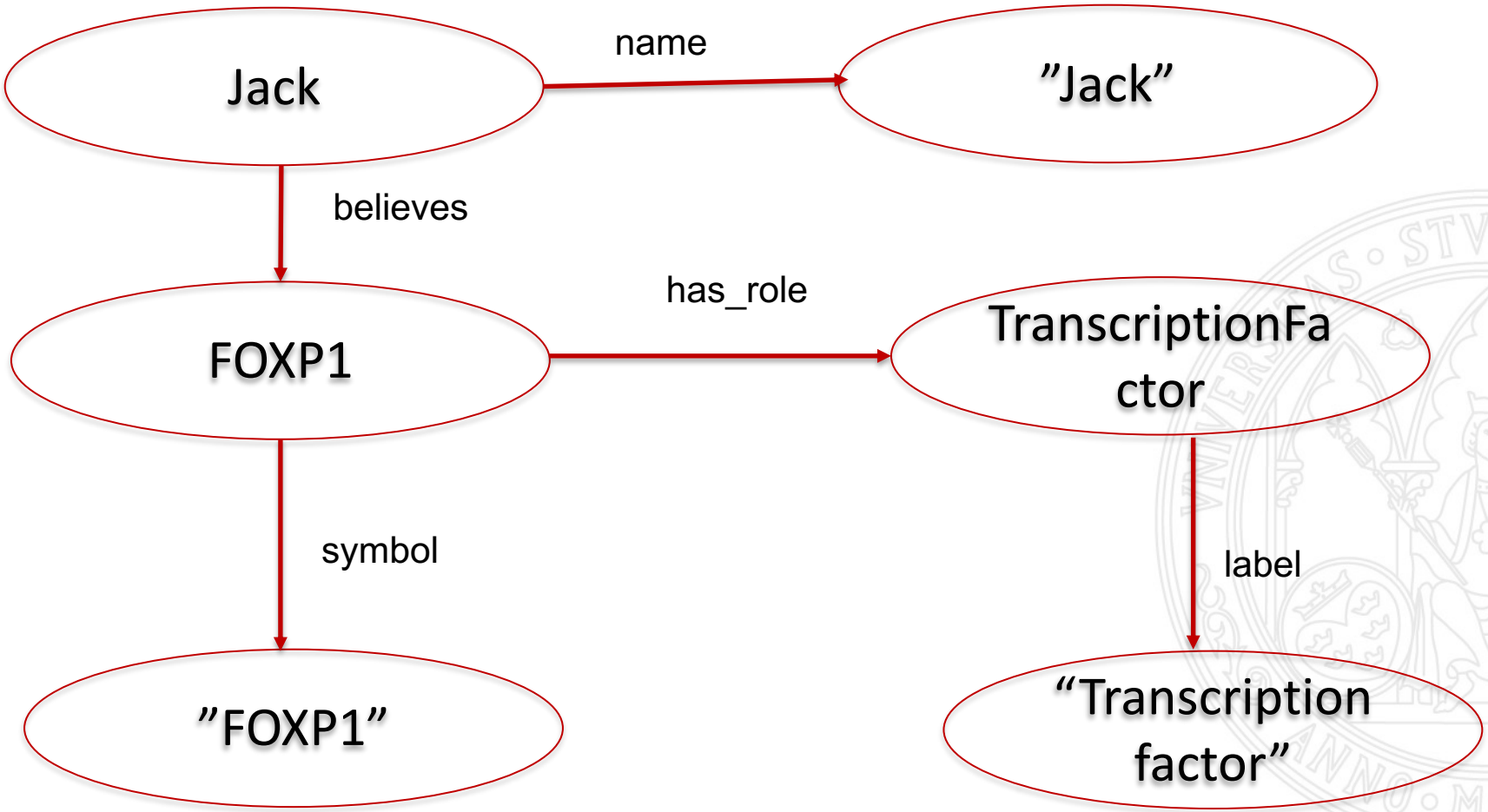
“FOXP1 plays the role of transcription factor”

UNIVERSIDAD DE MURCIA



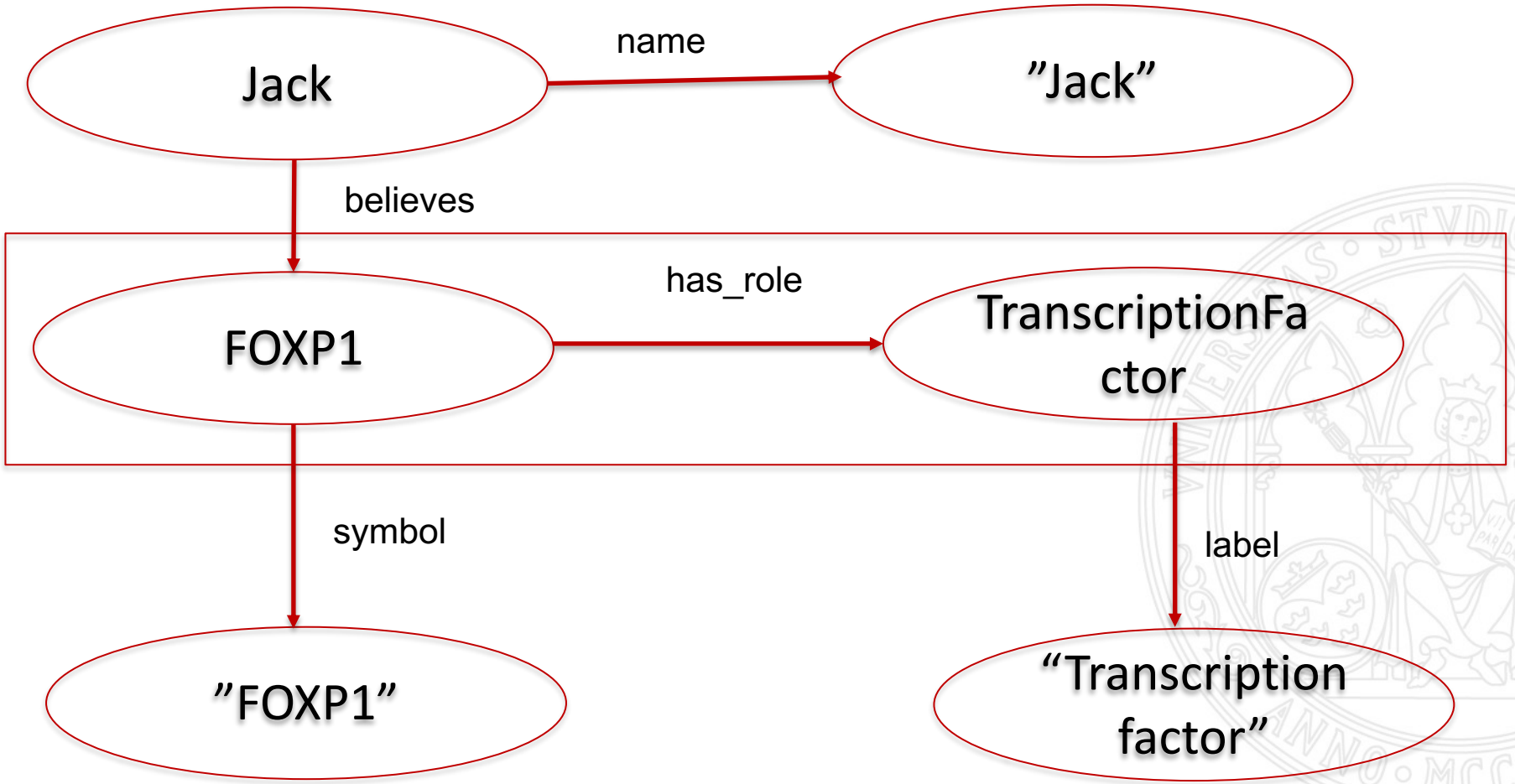
“Jack believes FOXP1 plays the role of transcription factor”

UNIVERSIDAD DE MURCIA



“Jack believes FOXP1 plays the role of transcription factor”

UNIVERSIDAD DE MURCIA

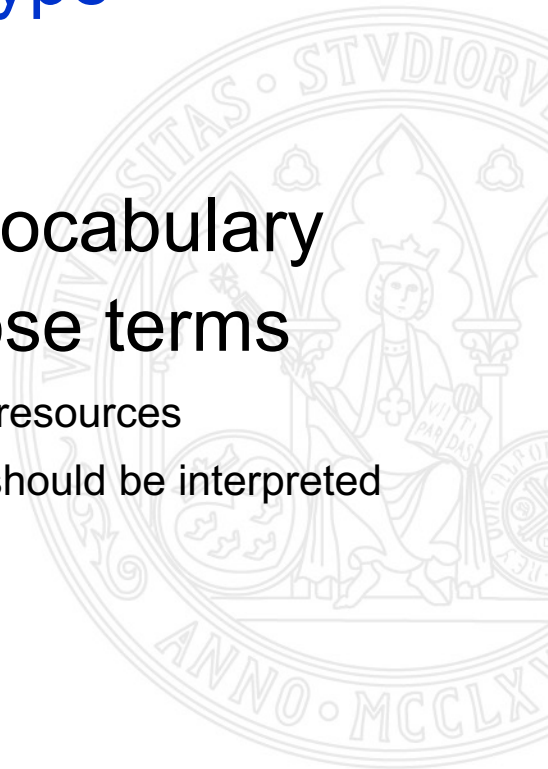


RDF Schema (RDFS)

UNIVERSIDAD DE MURCIA

- RDF does not give any special meaning to vocabulary such as subClassOf or type

- RDF Schema allows you to define vocabulary terms and the relations between those terms
  - it gives “extra meaning” to particular RDF predicates and resources
  - this “extra meaning”, or semantics, specifies how a term should be interpreted



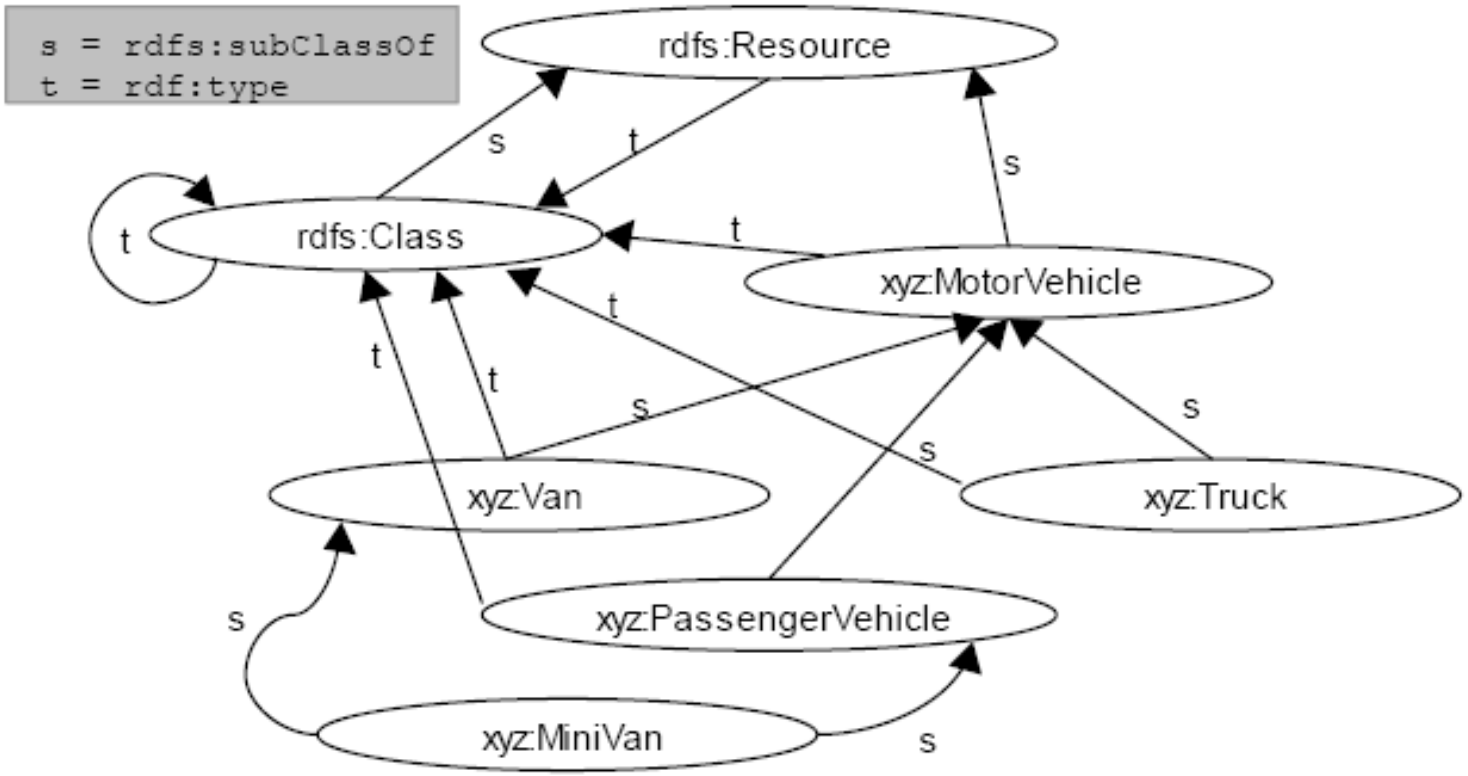
RDF Schema (RDFS) Examples

UNIVERSIDAD DE MURCIA

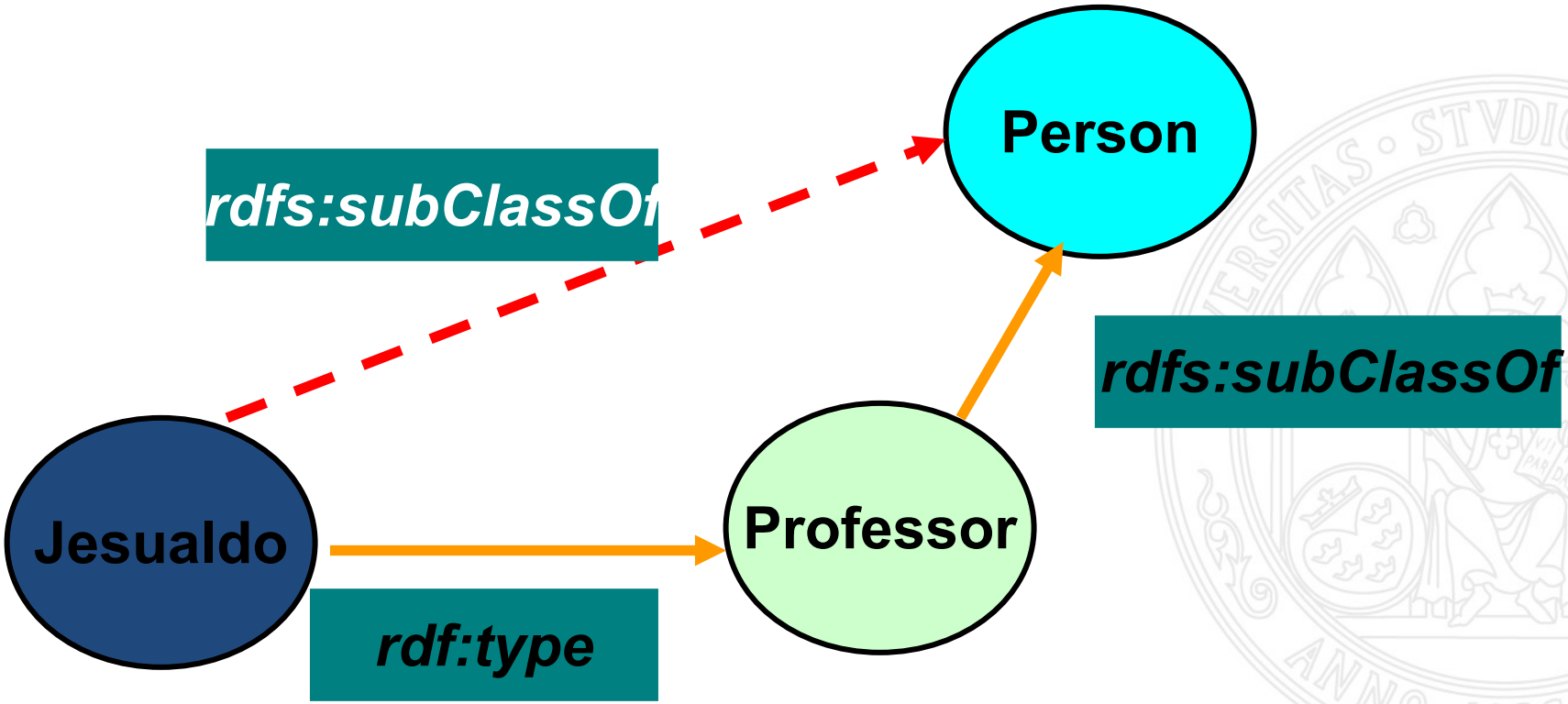
- RDF Schema terms (just a few examples):
  - Class
  - Property
  - type
  - subClassOf
  - range
  - domain
- These terms are the RDF Schema building blocks (constructors) used to create vocabularies:

```
<Person,type,Class>
<hasColleague,type,Property>
<Professor,subClassOf,Person>
<Rodrigo,type,Professor>
<hasColleague,range,Person>
<hasColleague,domain,Person>
```

RDF Schema (RDFS) Examples



Some inferences can be made now



Important properties

UNIVERSIDAD DE MURCIA

- rdf:type relates any resource to its Class
- Hierarchical properties
  - rdfs:subClassOf
  - rdfs:subPropertyOf
- rdfs: domain and rdfs: range
- Annotations: rdfs:seeAlso, rdfs:isDefinedBy

RDF/RDFS “Liberality”

UNIVERSIDAD DE MURCIA

- No distinction between classes and instances (individuals)

```
<Species,type,Class>
<Lion,type,Species>
<Leo,type,Lion>
```

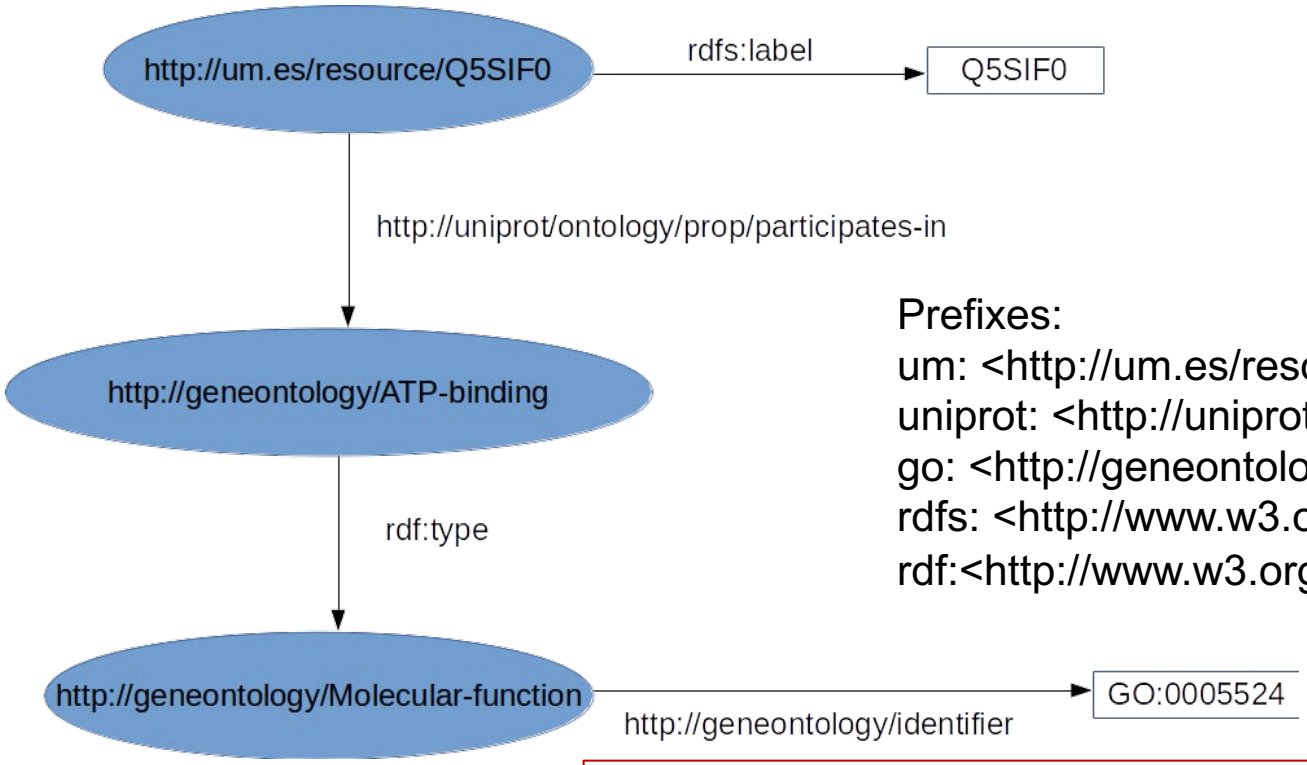
- Properties can themselves have properties

```
<hasDaughter,subPropertyOf,hasChild>
<hasDaughter,type,familyProperty>
```

Exercise 3

UNIVERSIDAD DE MURCIA

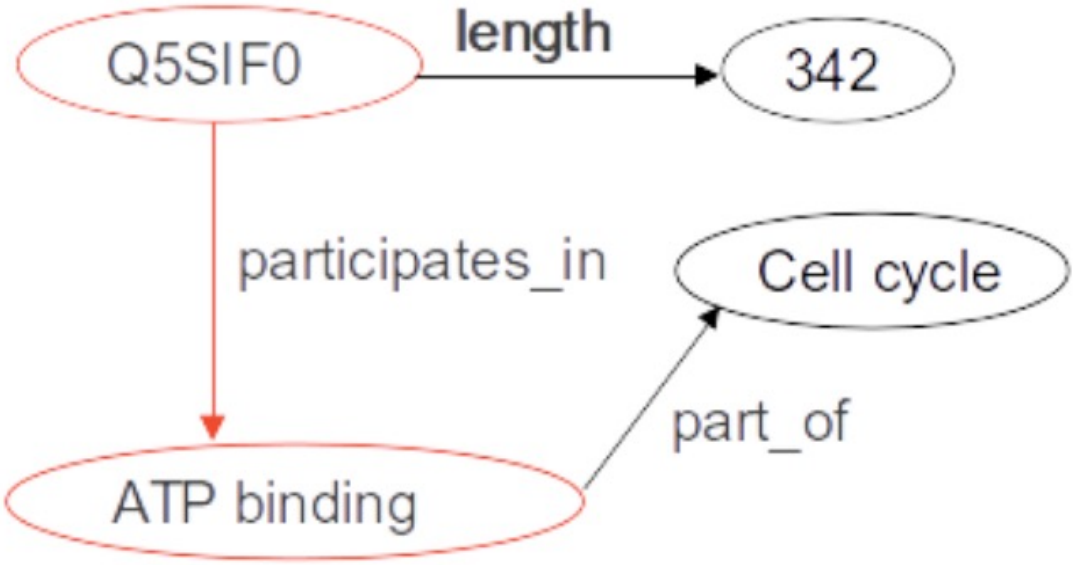
1) Write the following RDF model in Turtle (TTL)
2) Convert to RDF/XML (http://www.easyrdf.org/converter)
3) Validate and generate visualization (http://www.w3.org/RDF/Validator/)



Prefixes:
um: <http://um.es/resource/>
uniprot: <http://uniprot/ontology/prop/>
go: <http://geneontology/>
rdfs: <http://www.w3.org/2000/01/rdf-schema#>
rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>

Hint for properties: http://purl.uniprot.org/core/ /

Exercise 4: Same steps for this data



Revisiting exercises 4 and 5

UNIVERSIDAD DE MURCIA

- Now use the real URI of the GO TERMS

- Search them in Ontobee: http://www.ontobee.org/ontology/GO

Exercise 6: Modeling mammal data

UNIVERSIDAD DE MURCIA

order	family	Genus	species	mass(g)	gestation(mo)	newborn(g)	weaning(mo)	wean mass(g)	AFR(mo)	max. life(mo)	litter size	litters/year	refs
Artiodactyla	Antilocapridae	Antilocapra	americana	45375.00	8.13	3246.36	3.00	8900.00	13.53	142	1.85	1.00	"1,2,6,9,23,26,27"
Artiodactyla	Bovidae	Addax	nasomaculatus	182375.00	9.39	5480.00	6.50	-999.00	27.27	308	1.00	0.99	"1,2,17,23,26"
Artiodactyla	Bovidae	Aepyceros	melampus	41480.00	6.35	5093.00	5.63	15900.00	16.66	213	1.00	0.95	"1,2,8,9,23,29"
Artiodactyla	Bovidae	Alcelaphus	buselaphus	150000.00	7.90	10166.67	6.50	-999.00	23.02	240	1.00	-999.00	"1,2,17,23"
Artiodactyla	Bovidae	Ammodorcas	clarkei	28500.00	6.80	-999.00	-999.00	-999.00	-999.00	-999	1.00	-999.00	"1,2"
Artiodactyla	Bovidae	Ammotragus	lervia	55500.00	5.08	3810.00	4.00	-999.00	14.89	251	1.37	2.00	"1,2,9,11,17,23,29"
Artiodactyla	Bovidae	Antidorcas	marsupialis	30000.00	5.72	3910.00	4.04	-999.00	10.23	228	1.00	-999.00	"1,2,9,23,27"
Artiodactyla	Bovidae	Antilope	cervicapra	37500.00	5.50	3846.00	2.13	-999.00	20.13	255	1.00	1.89	"1,2,17"
Artiodactyla	Bovidae	Bison	bison	497666.67	8.93	20000.00	10.71	157500.00	29.45	300	1.00	1.00	"1,2,6,11,13,17,23"
Artiodactyla	Bovidae	Bison	bonasus	500000.00	9.14	23000.08	6.60	-999.00	29.99	324	1.00	1.00	"1,2,13,17,19"
Artiodactyla	Bovidae	Bos	grunniens	333000.00	8.88	18000.00	7.33	-999.00	24.27	300	1.00	0.75	"1,2,17,29"
Artiodactyla	Bovidae	Bos	frontalis	800000.00	9.02	23033.33	4.50	-999.00	24.16	314	1.17	0.89	"1,2,23"
Artiodactyla	Bovidae	Bos	javanicus	666666.67	9.83	-999.00	9.50	-999.00	25.54	319	1.33	1.00	"1,2,23"
Artiodactyla	Bovidae	Boselaphus	tragocamelus	169000.00	8.51	5875.00	-999.00	-999.00	29.97	260	1.51	1.13	"1,2,17,29"
Artiodactyla	Bovidae	Bubalus	depressicornis	-999.00	10.00	-999.00	-999.00	-999.00	-999.00	433	1.00	-999.00	"1,2,17"
Artiodactyla	Bovidae	Bubalus	mindorensis	233333.33	9.85	-999.00	-999.00	-999.00	-999.00	300	1.00	0.50	"11,13"
Artiodactvla	Bovidae	Bubalus	bubalis	950000.00	10.47	37500.00	7.50	-999.00	19.88	348	1.33	0.50	"1,2,15,29"

Exercise 6: Modeling text in RDF

UNIVERSIDAD DE MURCIA

Transcription is the process by which the information in a strand of DNA is copied into a new molecule of messenger RNA (mRNA). Transcription is carried out by an enzyme called RNA polymerase and a number of accessory proteins called transcription factors. Transcription factors can bind to specific DNA sequences called enhancer and promoter sequences in order to recruit RNA polymerase to an appropriate transcription site. Together, the transcription factors and RNA polymerase form a complex called the transcription initiation complex.

Questions, comments…

UNIVERSIDAD DE MURCIA