

UniProt SPARQL: A small tutorial



Swiss Institute of
Bioinformatics

UniProt

www.uniprot.org

UniProtKB Advanced Search

BLAST Align Retrieve/ID mapping Peptide search Help Contact

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

UniProtKB

UniProt Knowledgebase

Swiss-Prot (552,884) Manually annotated and reviewed.

TrEMBL (70,656,157) Automatically annotated and not reviewed.

UniRef Sequence clusters

UniParc Sequence archive

Proteomes

News

BLOG Twitter Facebook RSS

Forthcoming changes Planned changes for UniProt

UniProt release 2016_10 N-acyl amino acids: a new treatment for obesity? | Cross-references to DisGeNET and OpenTargets | Change of the cross-references to Phos...

UniProt release 2016_09 Ki-67: the great leap from simple marker to functional actor | Change of RDF representation of the cross...

News archive

Supporting data

Literature citations

Cross-ref. databases

Taxonomy

Diseases

Subcellular locations

Keywords

Getting started



Text search

Our basic text search allows you to search all the resources available

BLAST

Find regions of similarity between your sequences

Sequence alignments

Align two or more protein sequences using the Clustal Omega program

Retrieve/ID mapping

This tool merges the "Retrieve" and "ID Mapping" tools

Peptide search

UniProt data

Download latest release

Get the UniProt data

Statistics

View Swiss-Prot and TrEMBL statistics

How to cite us

The UniProt Consortium

Submit your data

Submit your sequences and annotation updates

SPARQL

Query UniProt data using a SQL like graph query language

Protein spotlight



Seeing Red

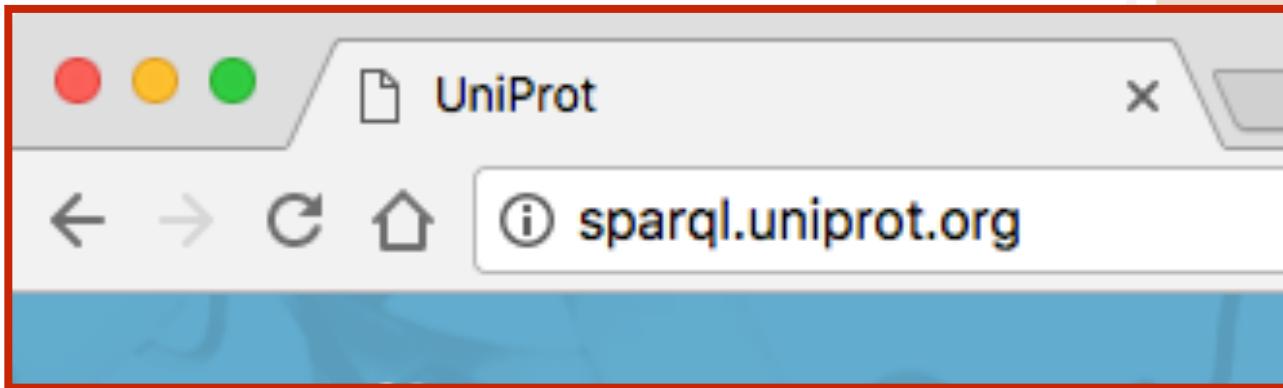
October 2016

The smallest of things can cause devastation. A word can hurt deeply. Distraction can create an accident. A spark can set off a fire. In the world of molecular biology, "small" rules. Life itself thrives on an intricate and very fragile harmony perpetuated by myriads of tiny molecules. And when something goes wrong, when a molecule changes its habits, it can create havoc. A few examples: cancer, Parkinson's disease, Alzheimer's and diabetes, to name but four...

Your SPARQL query

Add common prefixes

1



Submit Query

About

This SPARQL endpoint contains all UniProt data. It is free to access and supports the [SPARQL 1.1 Standard](#).

There are 23,181,093,815 triples in this release (2016_10). The query timeout is 45 minutes. All triples are available in the default graph. There are 17 named graphs.

Documentation

The documentation about UniProt RDF is spread into 2 parts

1. [Classes and predicates defined by the UniProt consortium](#)
2. [Statistics and diagrams](#)

Examples

1. Select all taxa from the [UniProt taxonomy](#): [\(show\)](#)
2. Select all bacterial taxa, and their scientific name, [UniProt taxonomy](#): [\(show\)](#)
3. Select all E-Coli K12 (including strains) [UniProt](#) and their amino acid sequence: [\(show\)](#)
4. Select the UniProt entry with the mnemonic 'IMAN': [\(show\)](#)
5. Select a mapping of UniProt to PDB entries using [Prot cross-references](#) to the [PDB database](#): [\(show\)](#)
6. Select all cross-references to external databases of the category '[3D structure databases](#)' of UniProt that are classified with the keyword '3Fe': [\(show\)](#)
7. Select all UniProt entries, and their recommended gene name, that have a preferred gene name that contains the text 'DNA': [\(show\)](#)
8. Select the preferred gene name and disease annotation of all human UniProt entries that are known to be involved in a disease: [\(show\)](#)
9. Select all human UniProt entries with a sequence variant that leads to a 'loss of function': [\(show\)](#)
10. Select all human UniProt entries with a sequence variant that leads to a tyrosine to phenylalanine substitution: [\(show\)](#)
11. Select all UniProt entries with annotated transmembrane regions and the regions' begin and end coordinates on the canonical sequence: [\(show\)](#)
12. Select all UniProt entries that were integrated on the 30th of November 2010: [\(show\)](#)
13. Was any UniProt entry integrated on the 9th of January 2013? [\(show\)](#)
14. Construct new triples of the type 'HumanProtein' from all human UniProt entries: [\(show\)](#)
15. Select all triples that relate to the EMBL CDS entry AA089367.1: [\(show\)](#)
16. Select all triples that relate to the taxon that describes *Homo sapiens*: [\(show\)](#)

News



Forthcoming changes

Planned changes for UniProt

N-acyl amino acids: a new treatment for obesity? | Cross-references to DisGeNET | Cross-references to OpenTargets
[UniProt release 2016_10](#)

Ki-67: the great leap from simple marker to functional actor | Change of RDF representation of the cross-references to family and domain databases | Change of RDF representation of the cross-references to EMBL
[UniProt release 2016_09](#)

Butterfly fashion: all they need is cortex | Cross-references to Conserved Domains Database | Change of the cross-

Local on your machine

http://localhost:9999/

The screenshot shows the SPARQL Playground interface running locally at <http://localhost:9999>. The interface has a header with tabs for SPARQL Playground, Poster, Explore, Data, Documentation, About, Resources, and a SIB logo.

Extract some data

Here is an example on how to get the first 10 rows of a dataset. Click on the examples on the right to continue your journey about learning SPARQL.

Show prefixes ... endpoint: <http://localhost:9999/sparql>

```
SELECT DISTINCT * WHERE {  
  ?s ?p ?o  
}  
LIMIT 10
```

When doing such a query it is important to set LIMIT 10.
This limit avoids performance issues, if the size of the dataset is un:

html ▾ Go Reset term finder

Tags ▾ Filter sparql examples

- 1) Select all taxa from the UniProt taxonomy [show diagram](#)
- 2) Select all taxa, and their scientific name, from the UniProt taxonomy [show diagram](#)
- 3) All Swiss-Prot entries [show diagram](#)
- 4) Select UniProt records with a related disease [show diagram](#)
- 5) Select UniProt records with a related disease and a comment [show diagram](#)
- 6) Select UniProt records with a related disease plus more information about the disease [show diagram](#)
- 7) Proteins that are in the genus Hominidae [show diagram](#)
- 8) Select Proteins similar to Human EPO isoform 3 [show diagram](#)
- 9) Describe an UniParc record [show diagram](#)

UniProtKB

Protein
aka
UniProt record

Taxonomy

UniRef

Cluster
aka
UniRef record

90%

100%

50%

UniParc

Sequence
aka
UniParc

UniProtKB

Protein
aka
UniProt record

Taxonomy

UniRef

Cluster
aka
UniRef record

UniParc

Sequence
aka
UniParc

50%

90%

100%

RHO - Rhodopsin - Homo sapiens

www.uniprot.org/uniprot/P08100

UniProtKB Advanced Help

BLAST Align Retrieve/ID mapping Peptide search

UniProtKB - P08100 (OPSD_HUMAN)

Display

Entry Publications Feature viewer Feature table

None

Function Names & Taxonomy Subcellular location Pathology & Biotech PTM / Processing Expression Interaction Structure Family & Domains Sequence Cross-references

Protein Rhodopsin
Gene RHO
Organism Homo sapiens (Human)

Status  Reviewed - Annotation score:  - Experimental evidence at protein level ⁱ

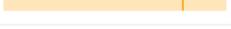
Function ⁱ

Photoreceptor required for image-forming vision at low light intensity. Required for photoreceptor cell viability after birth. Light-induced isomerization to all-trans retinal triggers a conformational change leading to G-protein activation and release of all-trans retinal.

Absorption ⁱ

Abs(max)=495 nm

Sites

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier
Metal binding ⁱ	201 – 201		1 Zinc  By similarity 		
Binding site ⁱ	265 – 265		1 Retinal chromophore  By similarity 		
Metal binding ⁱ	279 – 279		1 Zinc  By similarity 		

elixir © 2016 SIB SWITZERLAND

 Swiss Institute of Bioinformatics

RHO - Rhodopsin - Homo sapiens

www.uniprot.org/uniprot/P08100



UniProtKB ▾

LAST Align Retrieve/ID mapping Peptide search

UniProtKB - P08100 (OPSD_HUMAN)

display

BLAST

Align

Format

Add to basket

History

try

publications

feature viewer

feature table

Protein | **Rhodopsin**

Gene | **RHO**

Organism | *Homo sapiens (Human)*

RHO - Rhodopsin - Homo sapiens

www.uniprot.org/uniprot/P08100.ttl



UniProtKB ▾

LAST Align Retrieve/ID mapping Peptide search

UniProtKB - P08100 (OPSD_HUMAN)

display

BLAST

Align

Format

Add to basket

History

try

publications

feature viewer

feature table

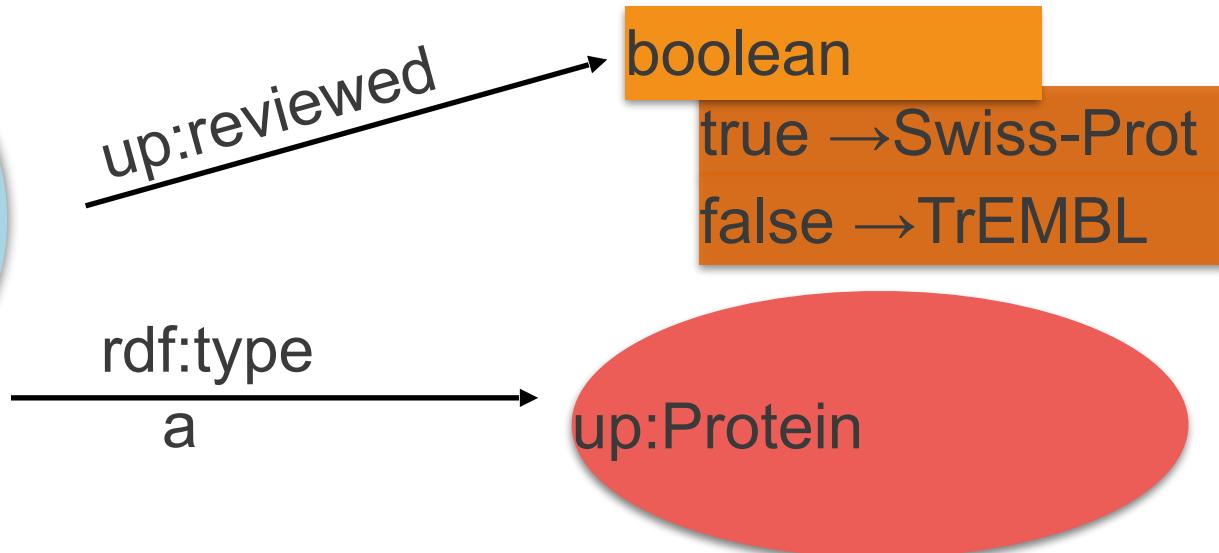
Protein | **Rhodopsin**

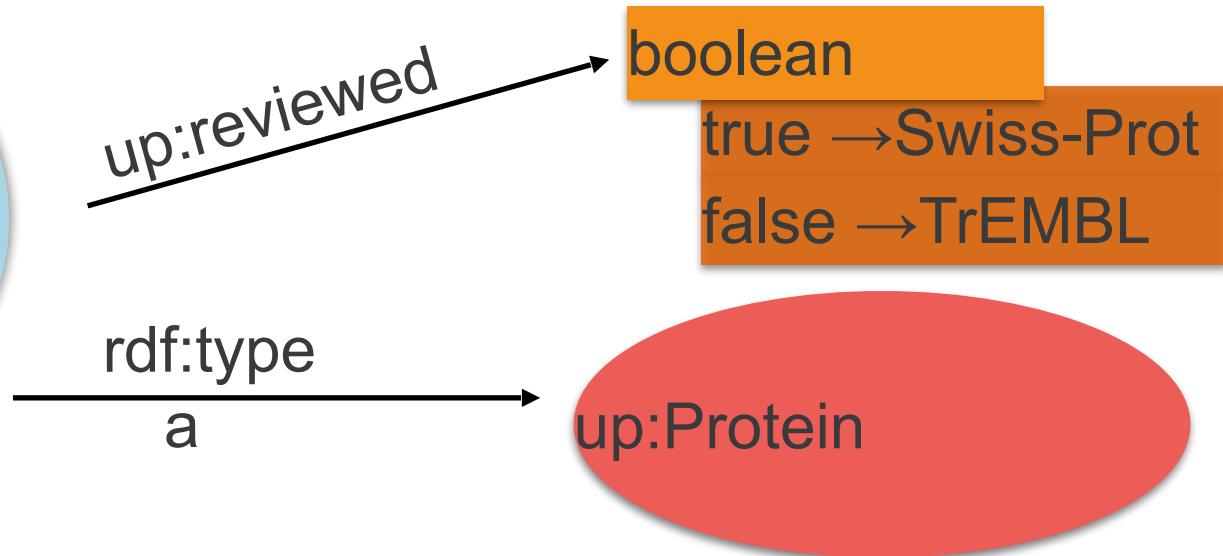
Gene | **RHO**

Organism | *Homo sapiens (Human)*

```
@prefix owl: <http://www.w3.org/2002/07/owl#> .  
@prefix position: <http://purl.uniprot.org/position/> .  
@prefix pubmed: <http://purl.uniprot.org/pubmed/> .  
@prefix range: <http://purl.uniprot.org/range/> .  
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .  
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .  
@prefix sha: <http://purl.uniprot.org/SHA-384/> .  
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .  
@prefix taxon: <http://purl.uniprot.org/taxonomy/> .  
@prefix tissue: <http://purl.uniprot.org/tissues/> .  
@prefix up: <http://purl.uniprot.org/core/> .  
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .  
<P08100> rdf:type up:Protein ;  
    up:reviewed true ;  
    up:created "1988-08-01"^^xsd:date ;  
    up:modified "2016-11-02"^^xsd:date ;  
    up:version 189 ;  
    up:mnemonic "OPSD_HUMAN" ;  
    up:replaces <Q16414> ,  
        <Q2M249> ;  
    up:citation citation:6589631 ,  
        sha:  
6E77DBF89E5E9DFD49581D074191102DC8B333C02C2F427703C1BB6  
AA0162D4903AE56ACFC602E27F399E51BA5371AC0 ,  
    citation:17974005 ,  
    citation:15489334 ,  
    citation:8566799 ,  
    citation:8045708 ,  
    citation:7987385 .
```

UniProtKB

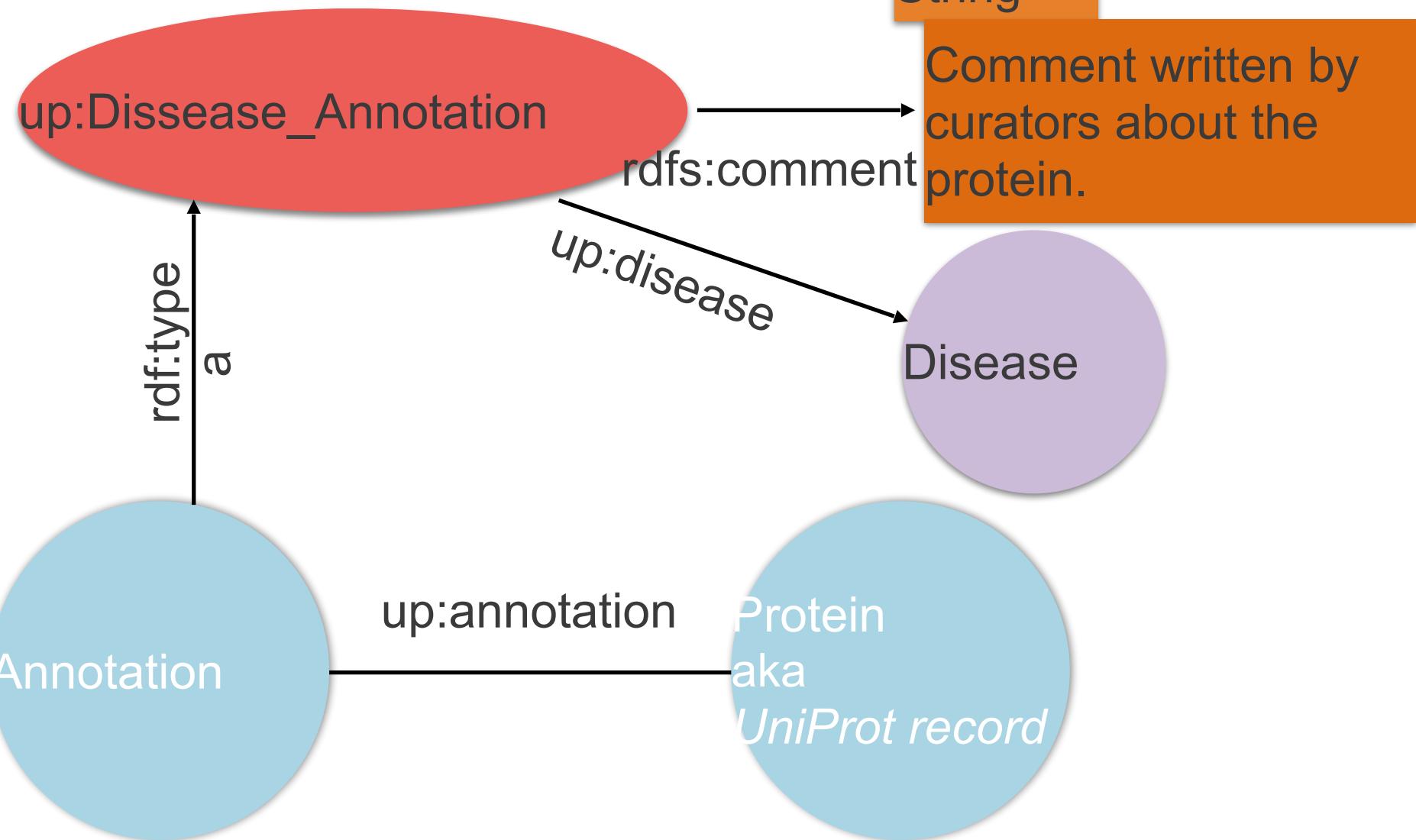


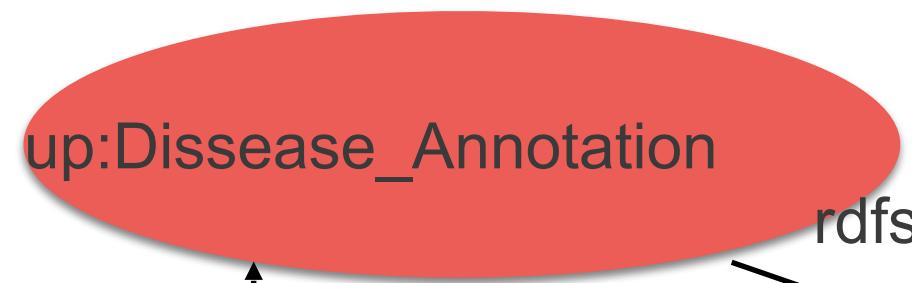


Swiss-Prot records: example 3

```
SELECT ?protein
WHERE {?protein a up:Protein;
          up:reviewed true}
```

UniProtKB





String

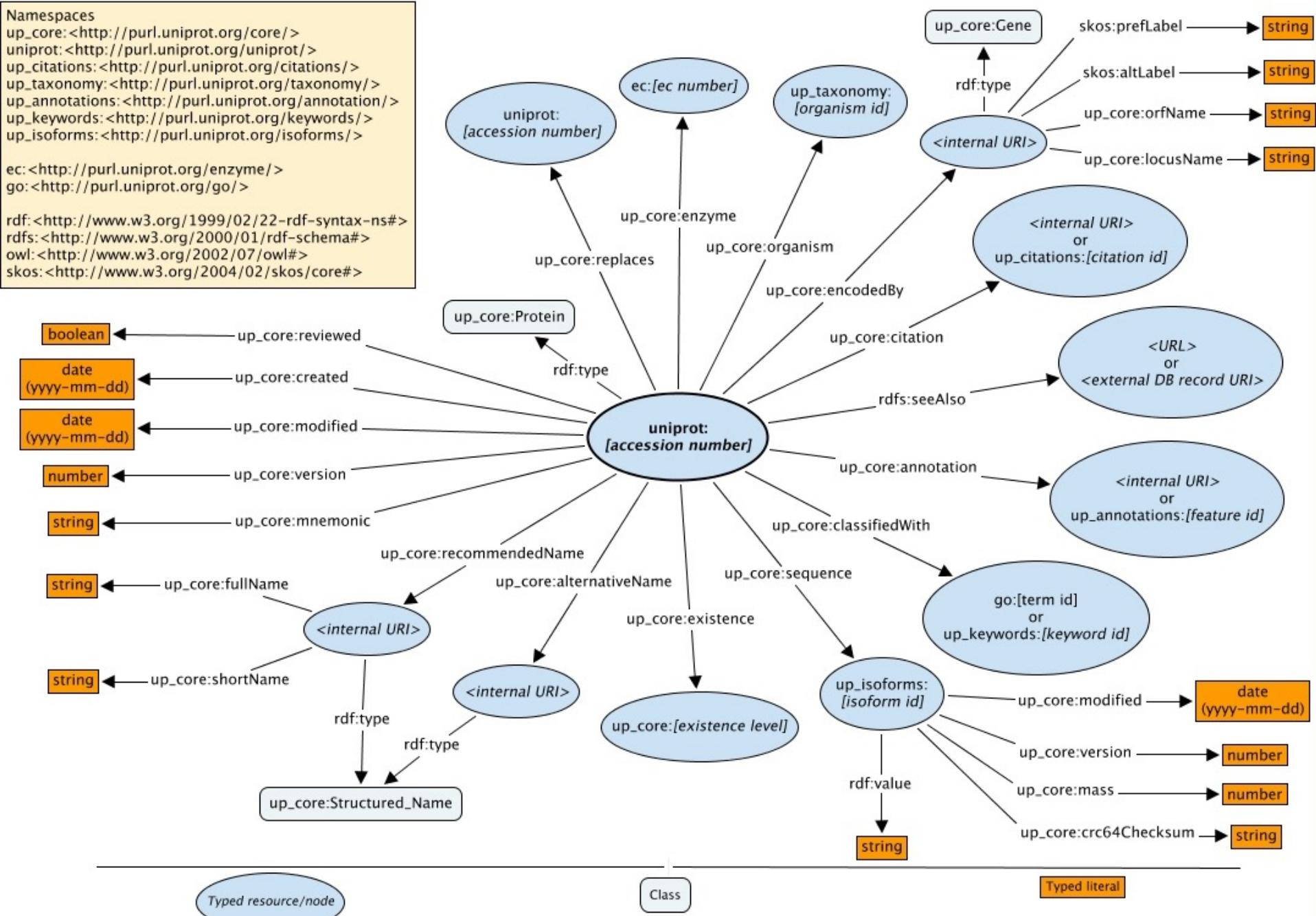
Comment written by
curators about the
protein.

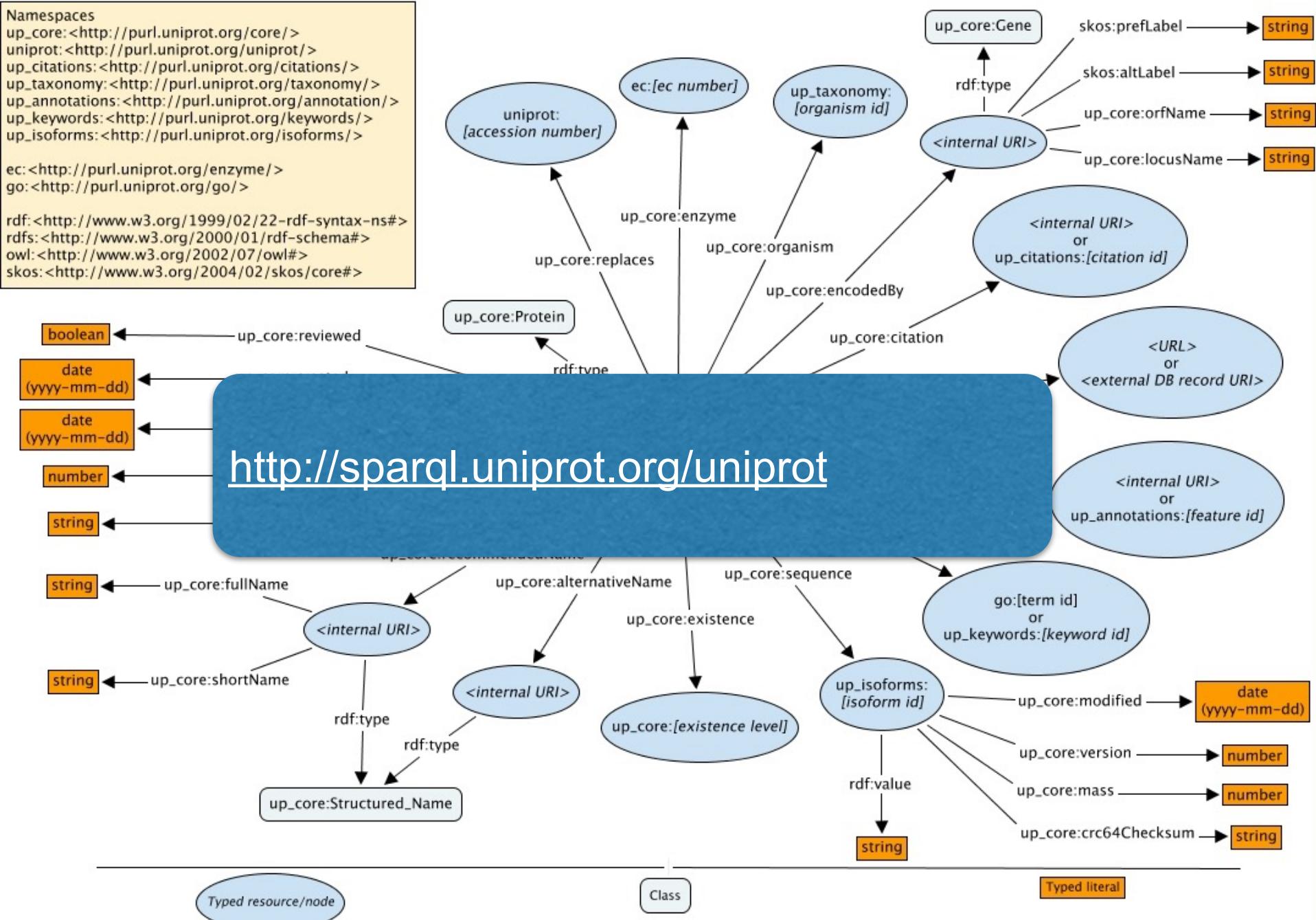
Disease

records with a disease: example4

```
SELECT ?protein ?disease
WHERE {?protein a up:Protein;
       up:annotation ?annotation .
       ?annotation up:disease ?disease}
```

Annota





UniProt RDF schema ontology x

www.uniprot.org/core/ AID

UniProt

BLAST Align Upload lists Help Contact

UniProt RDF schema ontology

Navigation panel

Classes (176)

up:Absorption_Annotation

Properties (165)

Object properties (77)

up:alternativeName

Datatype properties (88)

up:abstract

)
Properties (60)

up:API

Ontology description

<http://purl.uniprot.org/core/> (rdf:type owl:Ontology)

rdfs:label

UniProt RDF schema ontology

rdfs:comment

Properties and classes used for protein annotation. xsd:string

owl:versionInfo

2016_10 xsd:string

Classes

up:Absorption_Annotation (rdf:type owl:Class)

rdfs:subClassOf

up:Biophysicochemical_Annotation

rdfs:comment

Indicates the wavelength in nm at which photoreactive proteins such as opsins and DNA photolyases show maximal absorption. xsd:string

rdfs:label

Absorption xsd:string

up:Active_Site_Annotation (rdf:type owl:Class)

UniProtKB

Protein
aka
UniProt record

Taxonomy

UniRef

Cluster
aka
UniRef record

90%

100%

50%

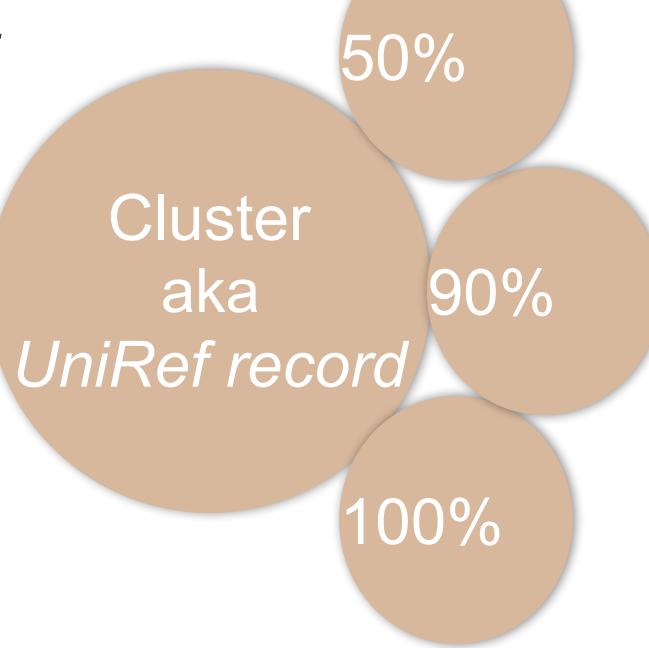
UniParc

Sequence
aka
UniParc

UniProtKB



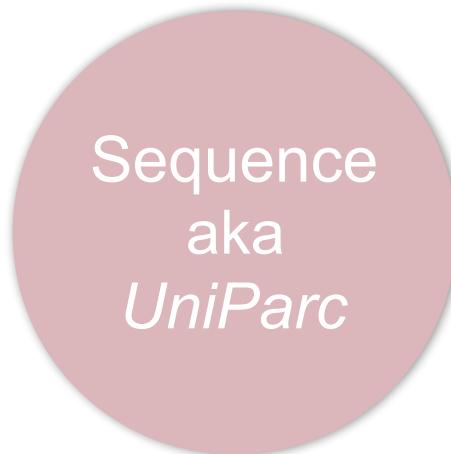
UniRef

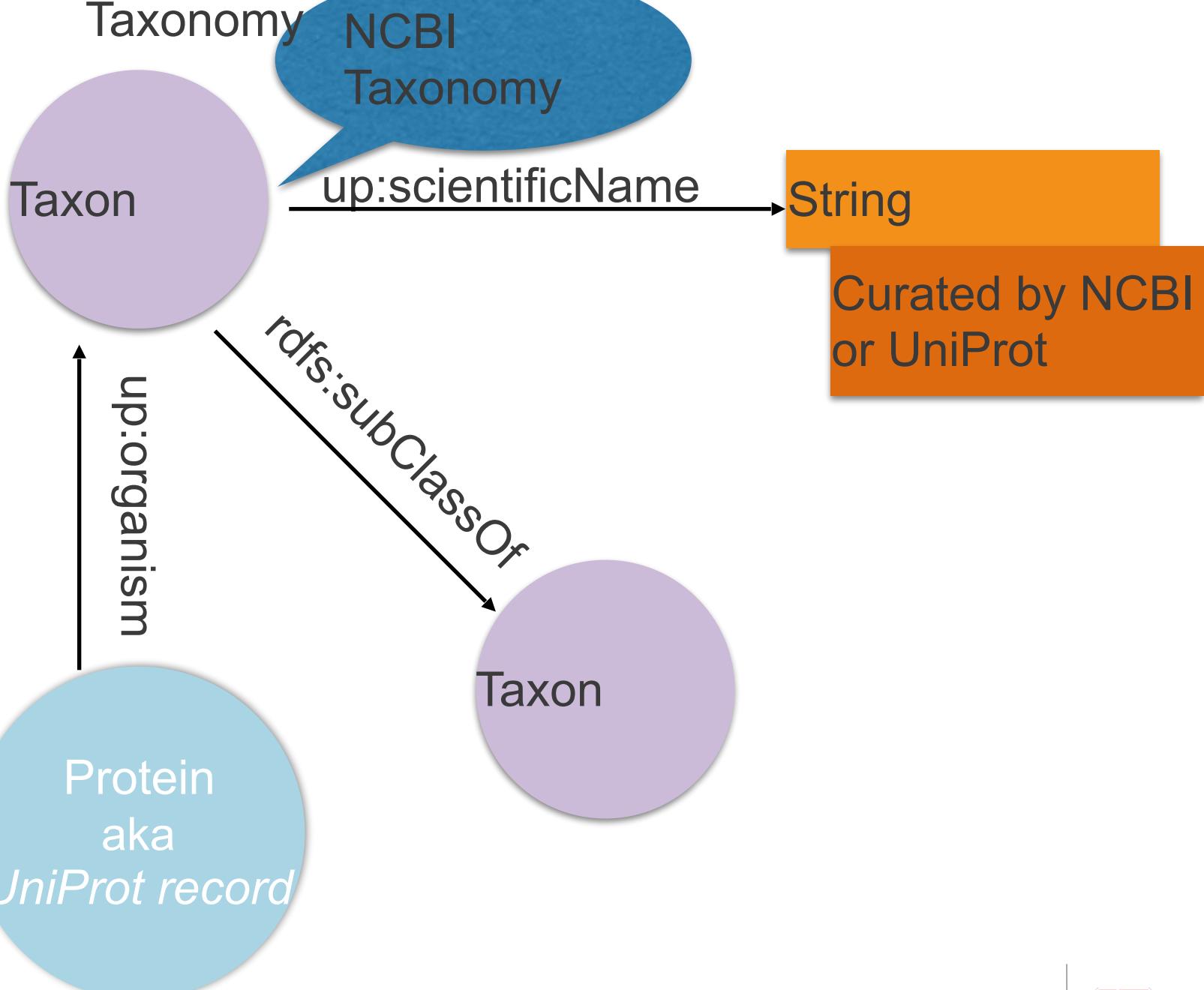


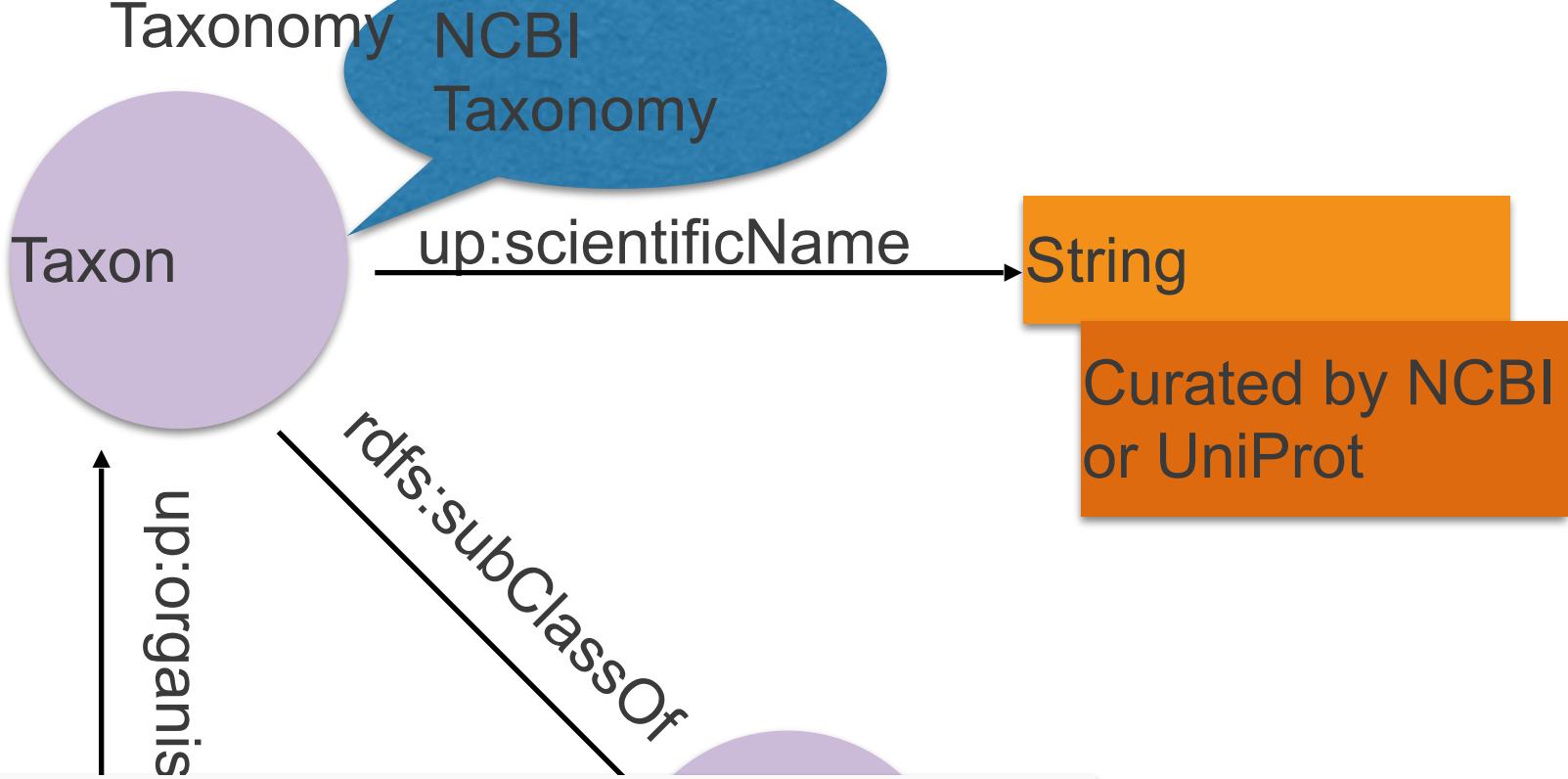
Taxonomy



UniParc







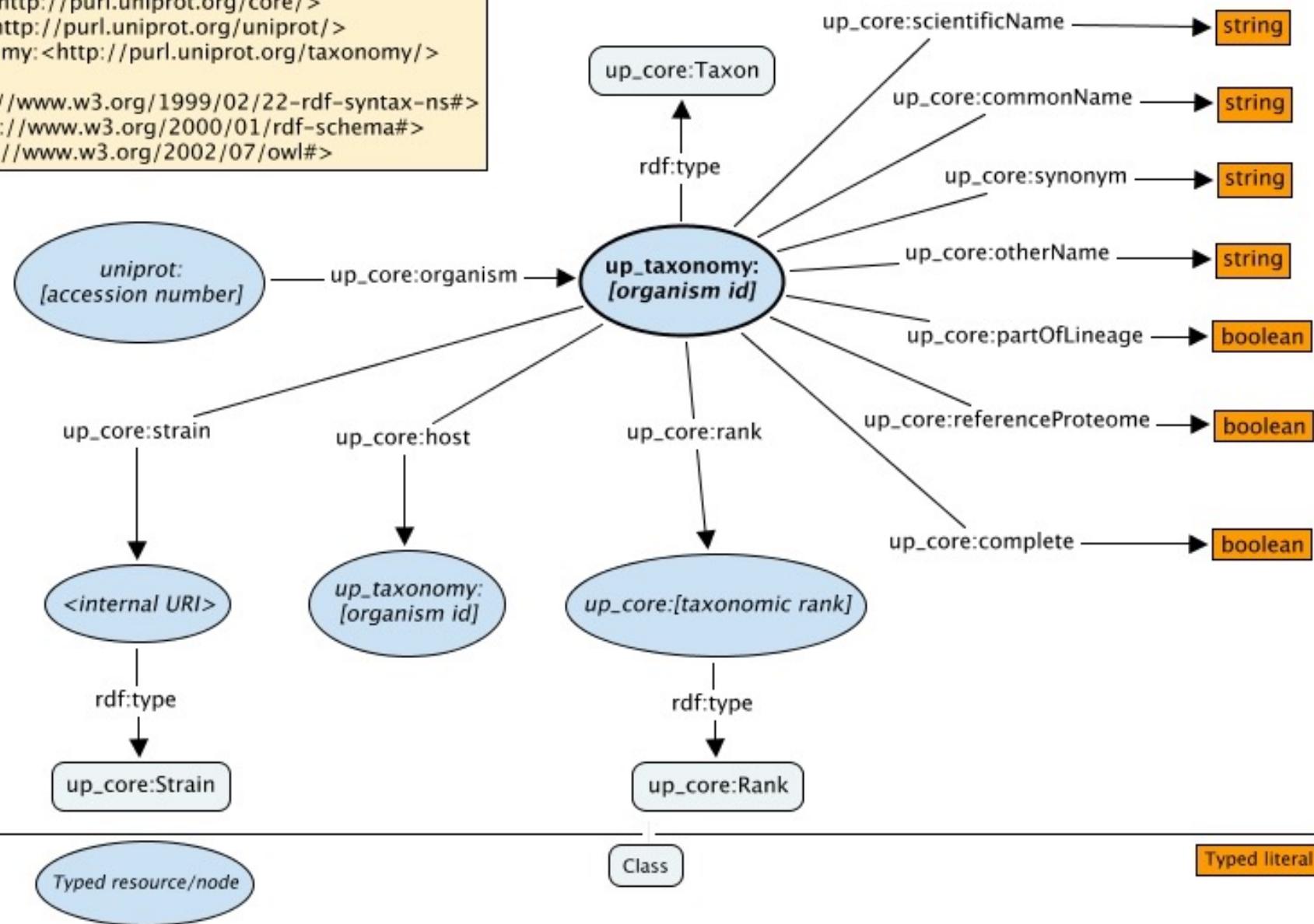
Proteins in the Homidea genus: example 5

```
SELECT ?protein ?organism
WHERE {?protein a up:Protein;
       up:organism ?organism .
       ?organism rdfs:subClassOf+ taxon:9604
}
```

Namespaces

up_core:<<http://purl.uniprot.org/core/>>
uniprot:<<http://purl.uniprot.org/uniprot/>>
up_taxonomy:<<http://purl.uniprot.org/taxonomy/>>

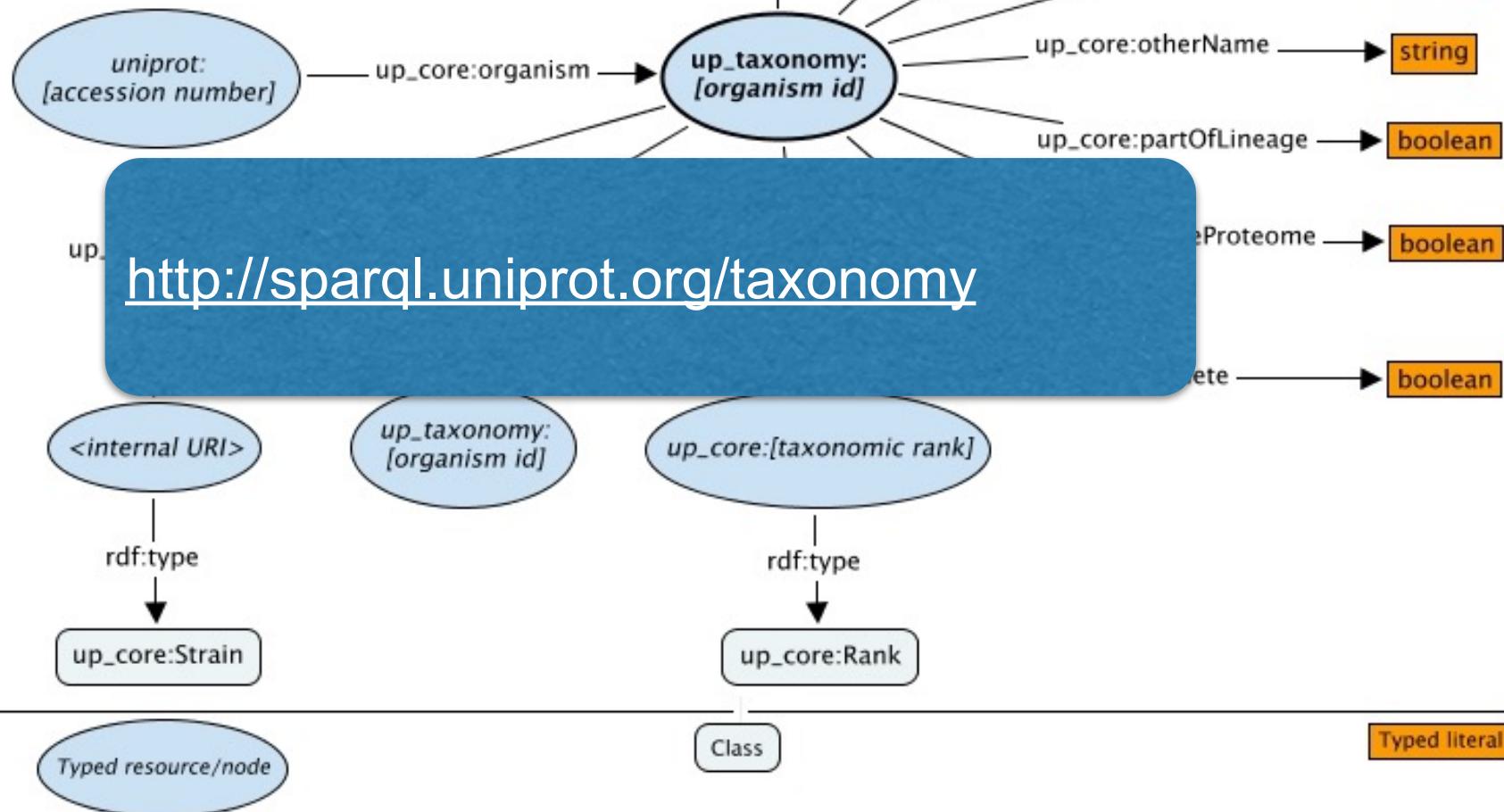
rdf:<<http://www.w3.org/1999/02/22-rdf-syntax-ns#>>
rdfs:<<http://www.w3.org/2000/01/rdf-schema#>>
owl:<<http://www.w3.org/2002/07/owl#>>



Namespaces

up_core:<<http://purl.uniprot.org/core/>>
uniprot:<<http://purl.uniprot.org/uniprot/>>
up_taxonomy:<<http://purl.uniprot.org/taxonomy/>>

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#>>



UniProtKB

Protein
aka
UniProt record

Taxonomy

UniRef

Cluster
aka
UniRef record

90%

100%

50%

UniParc

Sequence
aka
UniParc

UniProtKB

Protein
aka
UniProt record

Taxonomy

UniRef

Cluster
aka
UniRef record

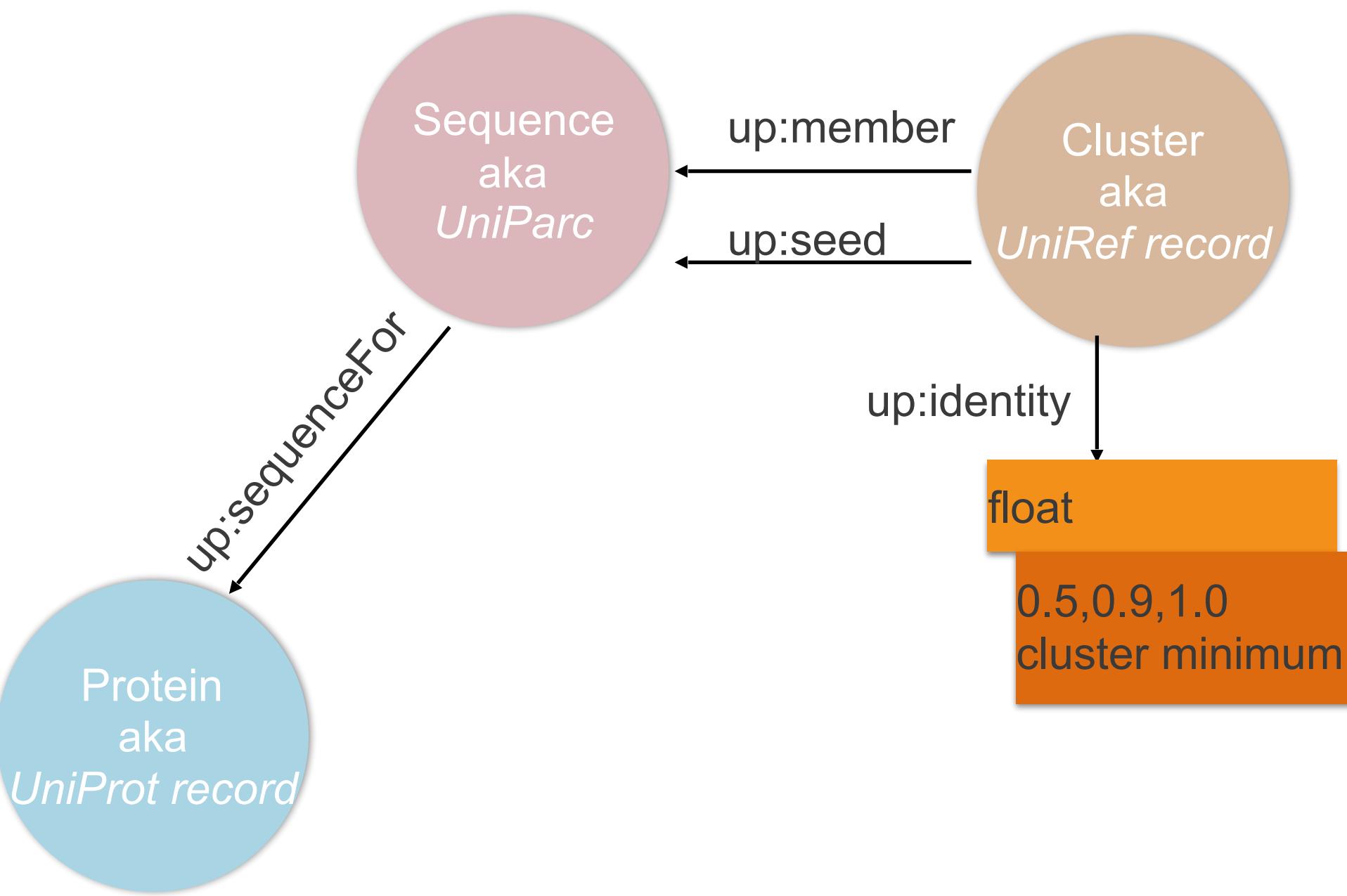
UniParc

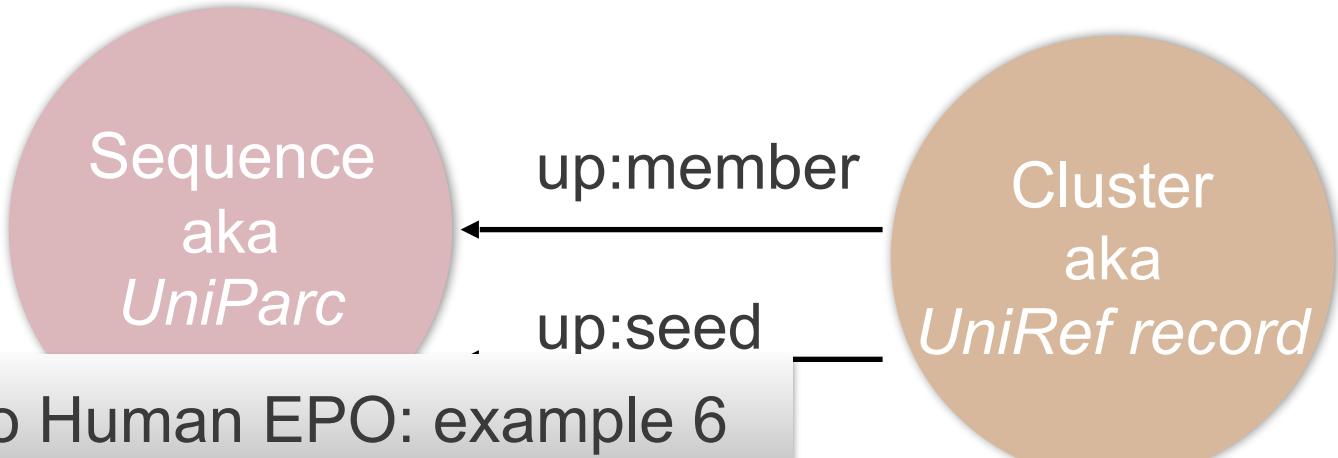
Sequence
aka
UniParc

50%

90%

100%





Protein similar to Human EPO: example 6

```
SELECT ?epoLike
WHERE {?cluster up:identity "0.9"^^xsd:float;
           up:member ?member,?member2 .
           ?member up:sequenceFor ?epoLike .
           ?member2 up:sequenceFor uniprot:P19235-3 .
           FILTER(?epoLike != uniprot:P19235-3)}
```

UniProtKB

Protein
aka
UniProt record

Taxonomy

UniRef

Cluster
aka
UniRef record

90%

100%

50%

UniParc

Sequence
aka
UniParc

UniProtKB

Protein
aka
UniProt record

Taxonomy

UniRef

Cluster
aka
UniRef record

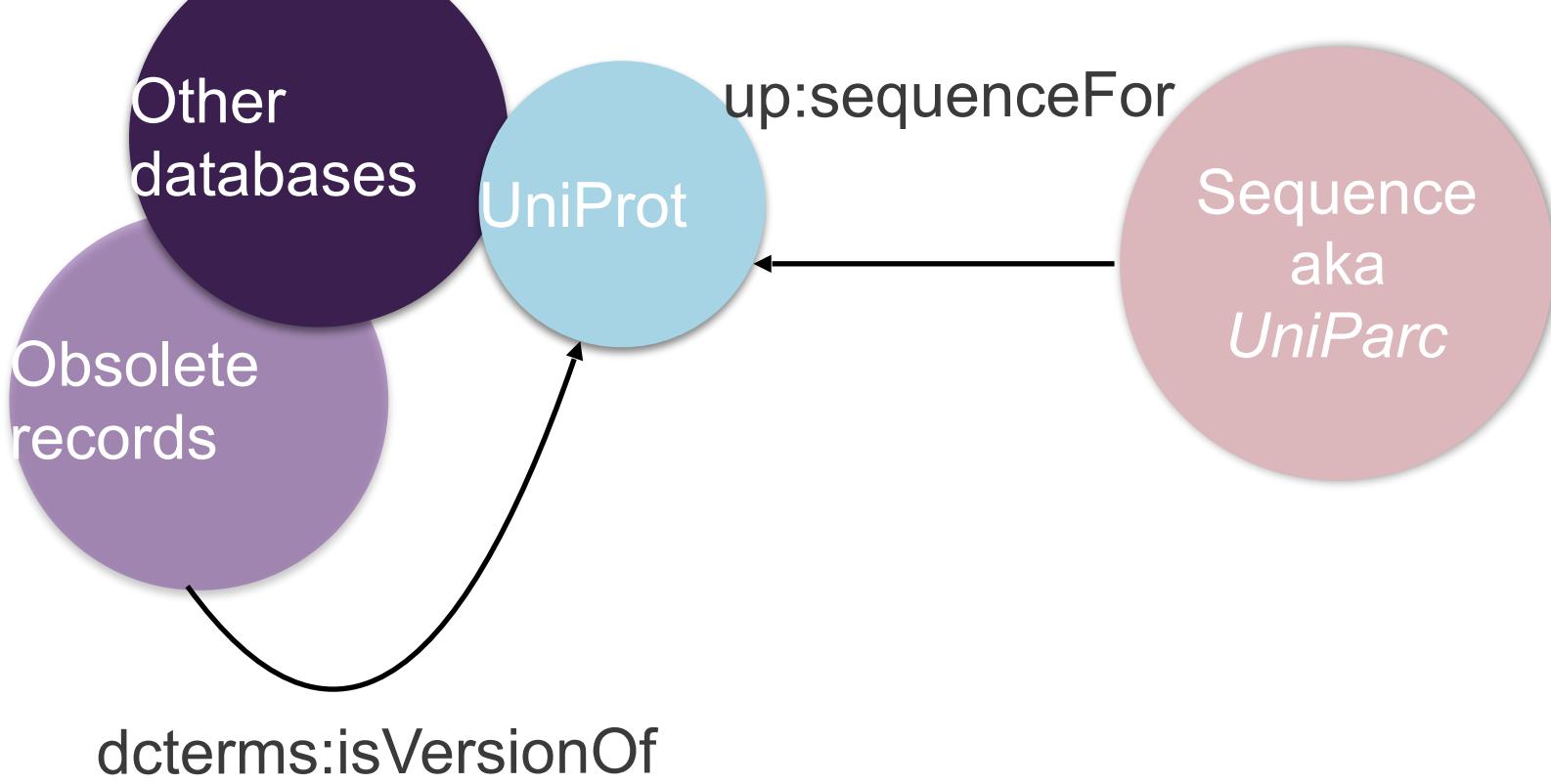
UniParc

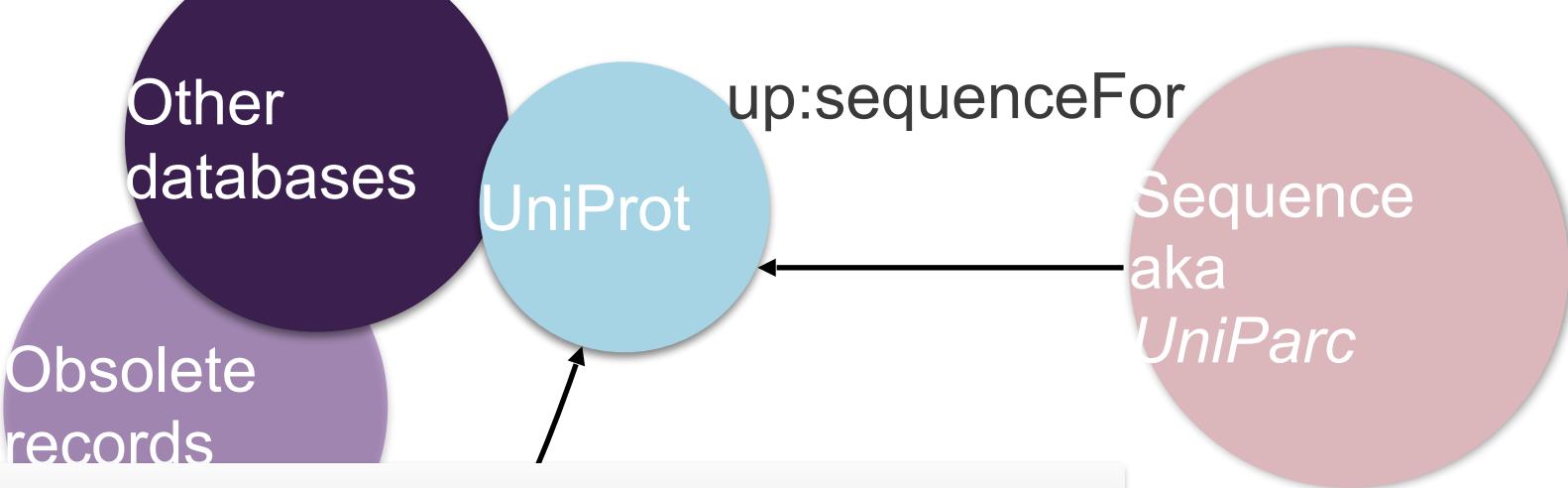
Sequence
aka
UniParc

50%

90%

100%





Find a uniparc protein via sequence identity
to an ENSEMBL one: example 7

```
SELECT ?sequence
WHERE {?sequence up:sequenceFor <http://
purl.uniprot.org/ENSEMBL/ENSP00000467809>.
}
```

Protein
aka
UniProt record

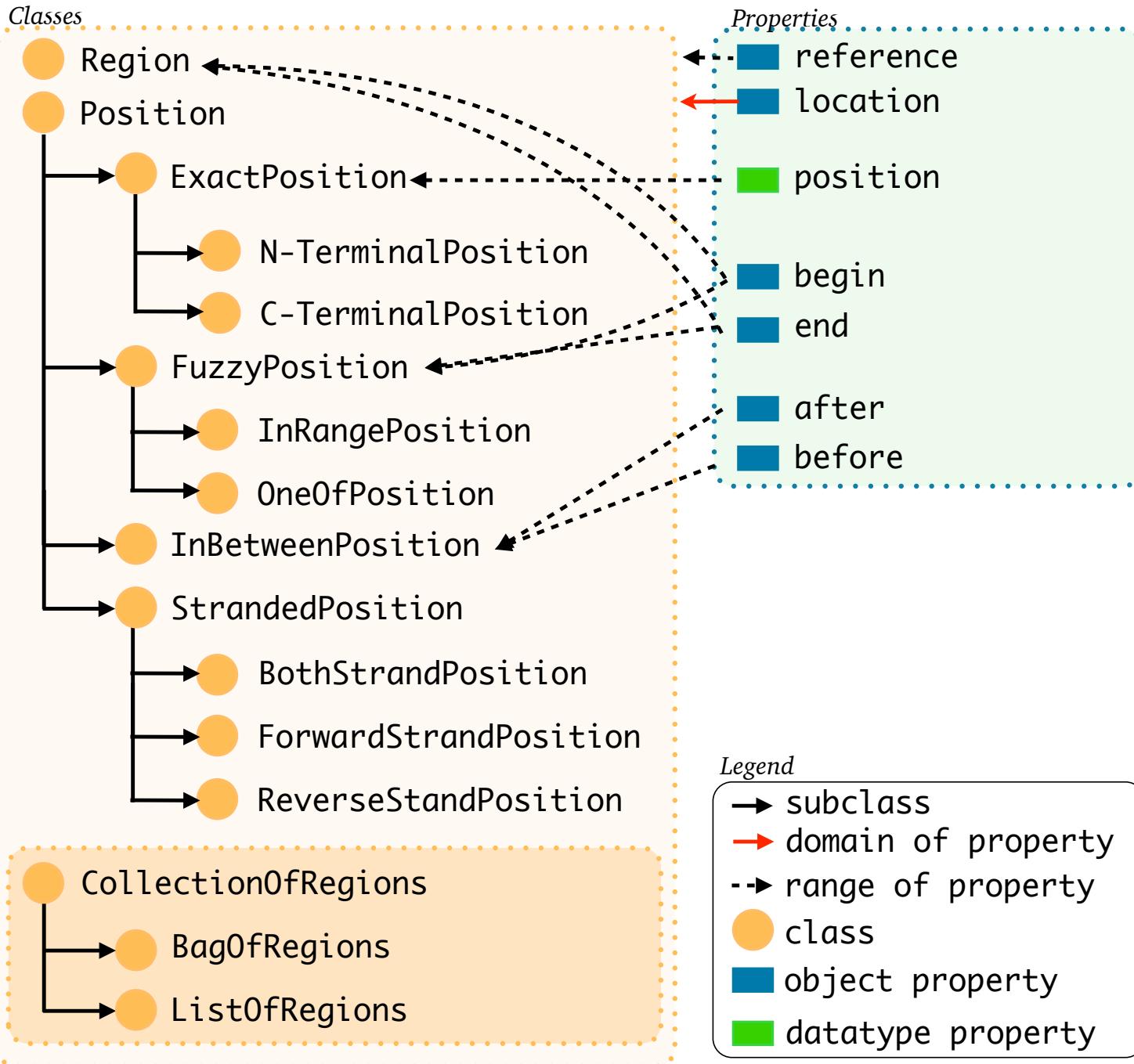
The diagram consists of two circles connected by a horizontal arrow. The left circle is light blue and contains the text 'Protein aka *UniProt record*'. The right circle is dark purple and contains the text 'Other databases'. A horizontal black arrow points from the left circle to the right circle. Above the arrow, the text 'rdfs:seeAlso' is written.

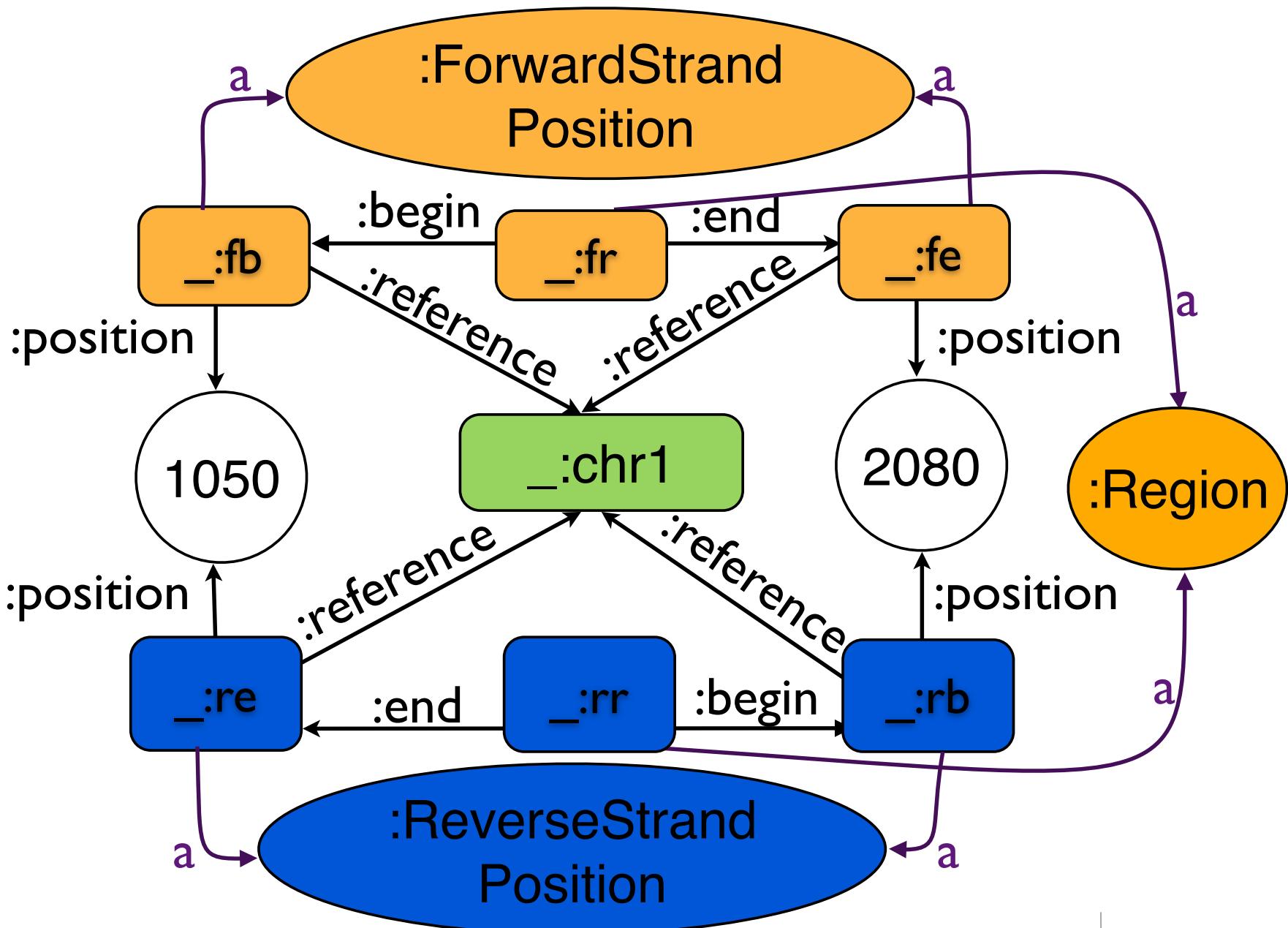
rdfs:seeAlso

Other
databases

Find a uniprot protein via crossreference to
an ENSEMBL one: example 8

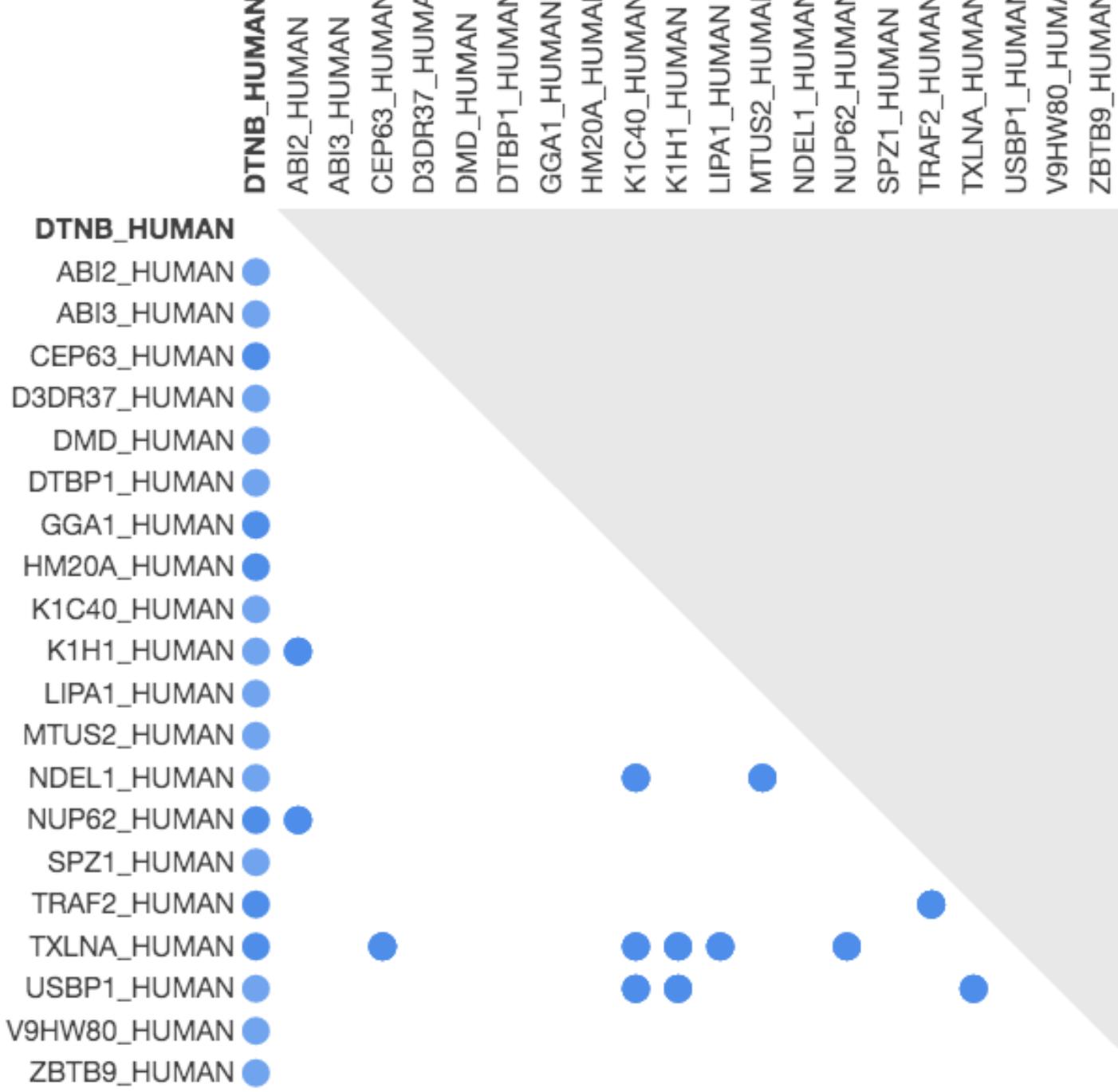
```
SELECT ?protein
WHERE {?protein rdfs:seeAlso
      <http://purl.uniprot.org/ENSEMBL/ENSP00000467809>.
}
```





Crazy things that are possible

```
PREFIX uniprotkb:<http://purl.uniprot.org/uniprot/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX up:<http://purl.uniprot.org/core/>
SELECT (SUM(?seq) AS ?totOfAA) ?aa
WHERE
{
  VALUES ?aa {"A" "R" "N" "D" "C" "Q" "E" "G" "H"
  "I" "L" "K" "M" "F" "P" "S" "T" "W" "Y" "V" "B" "Z" "J" "U" "O" "X" }
?protein up:organism taxon:1241371 .
?protein up:sequence/rdf:value ?sequence .
  BIND(CONCAT("[^",?aa,"]") as ?regex)
  BIND(STRLEN(REPLACE(?sequence, ?regex, "")) as ?seq)
} GROUP BY ?aa LIMIT 30
```



Complications, for standardisation

- FALDO: a semantic standard for describing the location of nucleotide and protein feature annotation
 - DOI: 10.1186/s13326-016-0067-z
- wwPDB RDF
- GO OBO2OWL
- Ensembl IRI
 - <http://www.uniprot.org/changes>



EMBL-EBI



W3C®

imI



DNA Data Bank of Japan

SPARQL



RDF

Linking

Open

SIB
swissprot
EBI



Semantic Web

Curation

UniProt

Reuse

Swiss
+
Lipids



Expertise



Swiss Institute of
Bioinformatics

Federated queries



Swiss Institute of
Bioinformatics

UniProt & Wikidata

<https://query.wikidata.org/>

```
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX up: <http://purl.uniprot.org/core/>
PREFIX taxon: <http://purl.uniprot.org/taxonomy/>
SELECT ?protein ?uniprot ?id WHERE {
?uniprot up:reviewed true ;
        up:organism taxon:9606 .
BIND (SUBSTR(STR(?uniprot),33) AS ?id)
#Convert IRI of UniProt to just the accession
SERVICE <http://query.wikidata.org/sparql>{
    ?protein wdt:P352 ?id .
}
```

BioModels & UniProt

<http://www.ebi.ac.uk/rdf/services/biomodels/sparql>

PREFIX rdf: <<http://www.w3.org/1999/02/22-rdf-syntax-ns#>>
PREFIX rdfs: <<http://www.w3.org/2000/01/rdf-schema#>>
PREFIX owl: <<http://www.w3.org/2002/07/owl#>>
PREFIX up: <<http://purl.uniprot.org/core/>>
PREFIX atlasterms: <<http://rdf.ebi.ac.uk/terms/atlas/>>

PREFIX sbmlrdf: <<http://identifiers.org/biomodels.vocabulary#>>

SELECT DISTINCT ?model ?uniprot ?taxonomy ?ec **WHERE** {

?model rdf:type sbmlrdf:SBMLModel .

?model ?linkstoellement ?modelement .

?idorgannot owl:sameAs ?uniprot.

?modelement ?qualifier ?idorgannot.

?modelement rdf:type ?elementType

FILTER(contains(str(?uniprot), "purl.uniprot.org/uniprot/"))

SERVICE<<http://sparql.uniprot.org/sparql>>{

?uniprot up:organism ?taxonomy .

?uniprot up:enzyme ?ec .}

UniProt and Ensembl

```
PREFIX core:<http://purl.uniprot.org/core/>
PREFIX uniprotkb:<http://purl.uniprot.org/uniprot/>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX obo: <http://purl.obolibrary.org/obo/>
PREFIX ensemblprotein: <http://rdf.ebi.ac.uk/resource/ensembl.protein/>
PREFIX ensemblterms: <http://rdf.ebi.ac.uk/terms/ensembl/>
PREFIX sio: <http://semanticscience.org/resource/>
```

```
SELECT ?protein ?transcript ?exon ?order {
  BIND(uniprotkb:P05067 as ?protein)
  ?protein rdfs:seeAlso ?dr .
  ?dr core:database <http://purl.uniprot.org/database/Ensembl> .
  BIND (iri(concat("http://rdf.ebi.ac.uk/resource/ensembl.transcript/", substr(str(?dr), 33))) as ?transcript)
  SERVICE <http://www.ebi.ac.uk/rdf/services/ensembl/sparql/>{
    ?transcript obo:SO_translates_to ?peptide .
    ?peptide a ensemblterms:protein .
    ?transcript obo:SO_has_part ?exon;
      sio:SIO_000974 ?orderedPart .
    ?orderedPart sio:SIO_000628 ?exon .
    ?orderedPart sio:SIO_000300 ?order .
  }
}
```

UniProt and Ensembl from 2016_11

```
PREFIX core:<http://purl.uniprot.org/core/>
PREFIX uniprotkb:<http://purl.uniprot.org/uniprot/>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX obo: <http://purl.obolibrary.org/obo/>
PREFIX ensemblprotein: <http://rdf.ebi.ac.uk/resource/ensembl.protein/>
PREFIX ensemblterms: <http://rdf.ebi.ac.uk/terms/ensembl/>
PREFIX sio: <http://semanticscience.org/resource/>
```

```
SELECT ?protein ?transcript ?exon ?order {
  BIND(uniprotkb:P05067 as ?protein)
  ?protein rdfs:seeAlso ?transcript .
  ?transcript core:database <http://purl.uniprot.org/database/Ensembl> .
  SERVICE <http://www.ebi.ac.uk/rdf/services/ensembl/sparql/>{
    ?transcript obo:SO_translates_to ?peptide .
    ?peptide a ensemblterms:protein .
    ?transcript obo:SO_has_part ?exon;
      sio:SIO_000974 ?orderedPart .
    ?orderedPart sio:SIO_000628 ?exon .
    ?orderedPart sio:SIO_000300 ?order .
  }
}
```

neXtProt to UniProt

snorql.nextprot.org

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

PREFIX taxon:<http://purl.uniprot.org/taxonomy/>

SELECT DISTINCT ?entry where {

?entry :isoform ?iso.

{

SERVICE <http://sparql.uniprot.org/sparql> {

SELECT DISTINCT ?viralinteractor **WHERE** # get viral proteins with an IntAct xref

{

 ?viralinteractor a up:Protein .

 ?viralinteractor rdfs:seeAlso/up:database <http://purl.uniprot.org/database/IntAct> .

 ?viralinteractor up:organism/rdfs:subClassOf/rdfs:subClassOf taxon:10239 .

}

}

?entry :isoform / :binaryInteraction / :interactant ?interactant. # NeXtprot entries with an IntAct binary interaction

?interactant skos:exactMatch ?viralinteractor . # interactant must be in the uniprot service result set to select the entry

}

UNION

{

SERVICE <http://sparql.uniprot.org/sparql> {

SELECT DISTINCT ?humprotein **WHERE** # get human proteins that share a PDB xref with a viral protein (same PDB id)

{

 ?humprotein a up:Protein .

 ?humprotein up:organism taxon:9606 .

 ?humprotein rdfs:seeAlso/up:database <http://purl.uniprot.org/database/PDB> .

 ?viralprotein a up:Protein ;

 rdfs:seeAlso ?db ;

 up:organism/rdfs:subClassOf/rdfs:subClassOf taxon:10239 .

}

}

BIND (IRI(CONCAT("http://nextprot.org/rdf/entry/NX_",substr(str(?humprotein),33,6))) as ?entry) # cast result to NeXtprot entry

}

DisGeNET and UniProt

<http://rdf.disgenet.org/sparql/>

```
SELECT ?protein
      ?comment
WHERE {
  ?protein a ncit:C17021; skos:exactMatch ?uniprot .
  FILTER(strstarts(str(?uniprot), "http://purl.uniprot.org/
uniprot"))
# Query UniProt for proteins with disease annotation
SERVICE <http://sparql.uniprot.org/sparql> {
  ?uniprot up:annotation ?annotation .
  ?annotation a up:Disease_Annotation ;
    rdfs:comment ?comment .
}
LIMIT 10
```

ChEMBL UniProt

PREFIX cco: <<http://rdf.ebi.ac.uk/terms/chembl#>>

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

PREFIX chembl_molecule: <<http://rdf.ebi.ac.uk/resource/chembl/molecule/>>

SELECT ?activity ?assay ?target ?targetcmpt ?uniprot ?fullName

WHERE {

?activity a cco:Activity ;

cco:hasMolecule chembl_molecule:CHEMBL941 ;

cco:hasAssay ?assay .

?assay cco:hasTarget ?target .

?target cco:hasTargetComponent ?targetcmpt .

?targetcmpt cco:targetCmptXref ?uniprot .

?uniprot a cco:UniprotRef

SERVICE <<http://sparql.uniprot.org/sparql/>>{

?uniprot up:recommendedName ?name .

?name up:fullName ?fullName .

}

Data modeling guidelines



Swiss Institute of
Bioinformatics

Guidelines for modelling

1. Model the data that you have
2. Other peoples data other peoples models
3. Makes sentences
4. Draw with pen and paper
5. Set or Instance

Model the data that you have

Biology is complicated

Model measurements not “facts”

Try to make statements that remain true

e.g. I eat lunch

e.g. On 2016/10/08 I ate lunch

1. Other peoples data other peoples models

GO (Gene Ontology) uses First Order Logic with an open world assumption to model GO using OWL.

Complicated, makes my head hurt. Yet if user knows GO they can use it in UniProt as well

Makes sentences

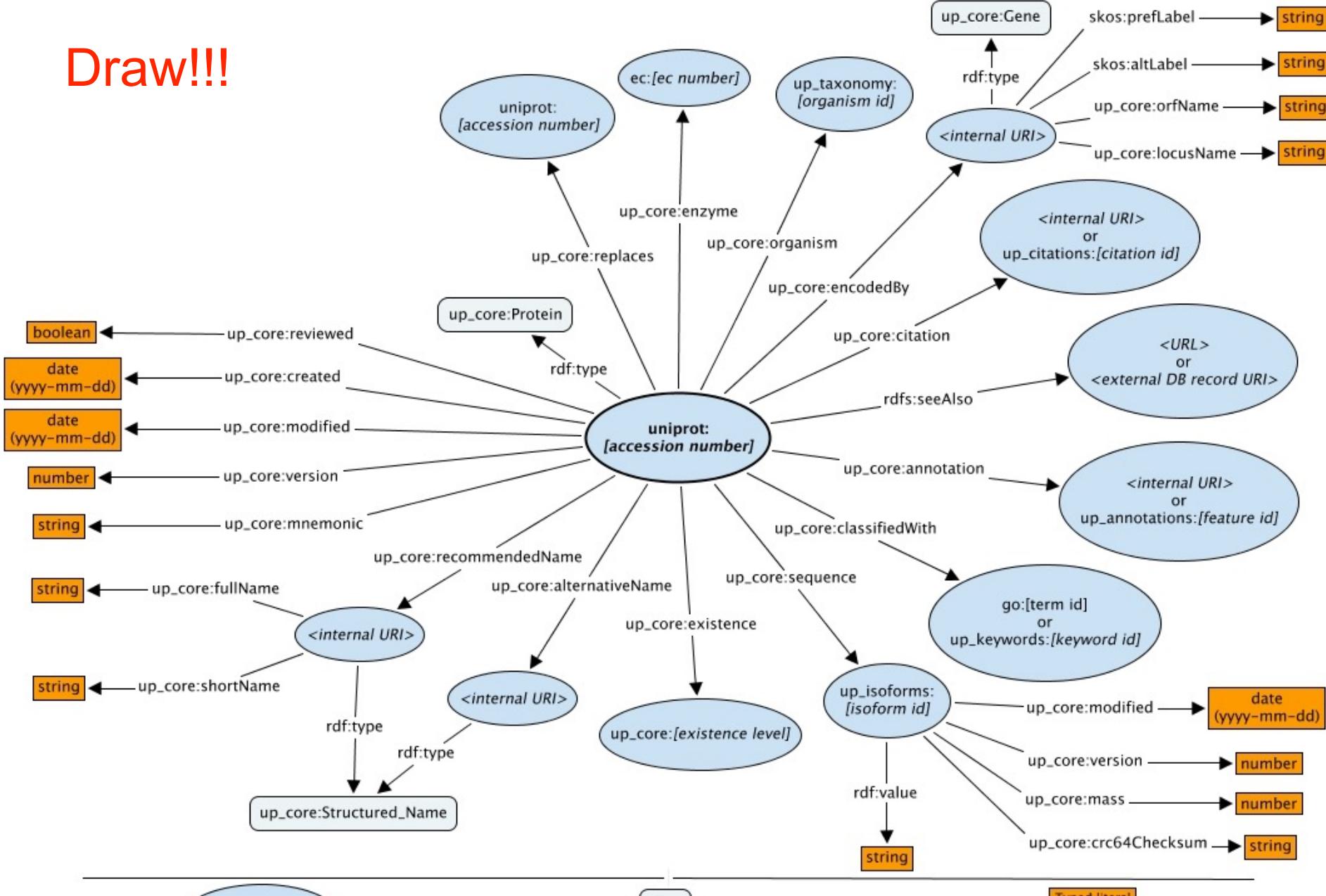
This UniProt entry reviewed, yes

<P05067> up:reviewed “true” .

Protein a interacts with Protein b

<P05067> up:interacts <P12345> .

Draw!!!



Set or Instance

rdf:type/owl:Class

me

you

your pet cat

rdfs:subClassOf/owl:Class

humanity

cats

reactions ← unless you see one set of molecules react

protein ← unless you can catch one molecule

Be Honest !!!

- Uncertainty
- Arbitrary
- Lack of identity