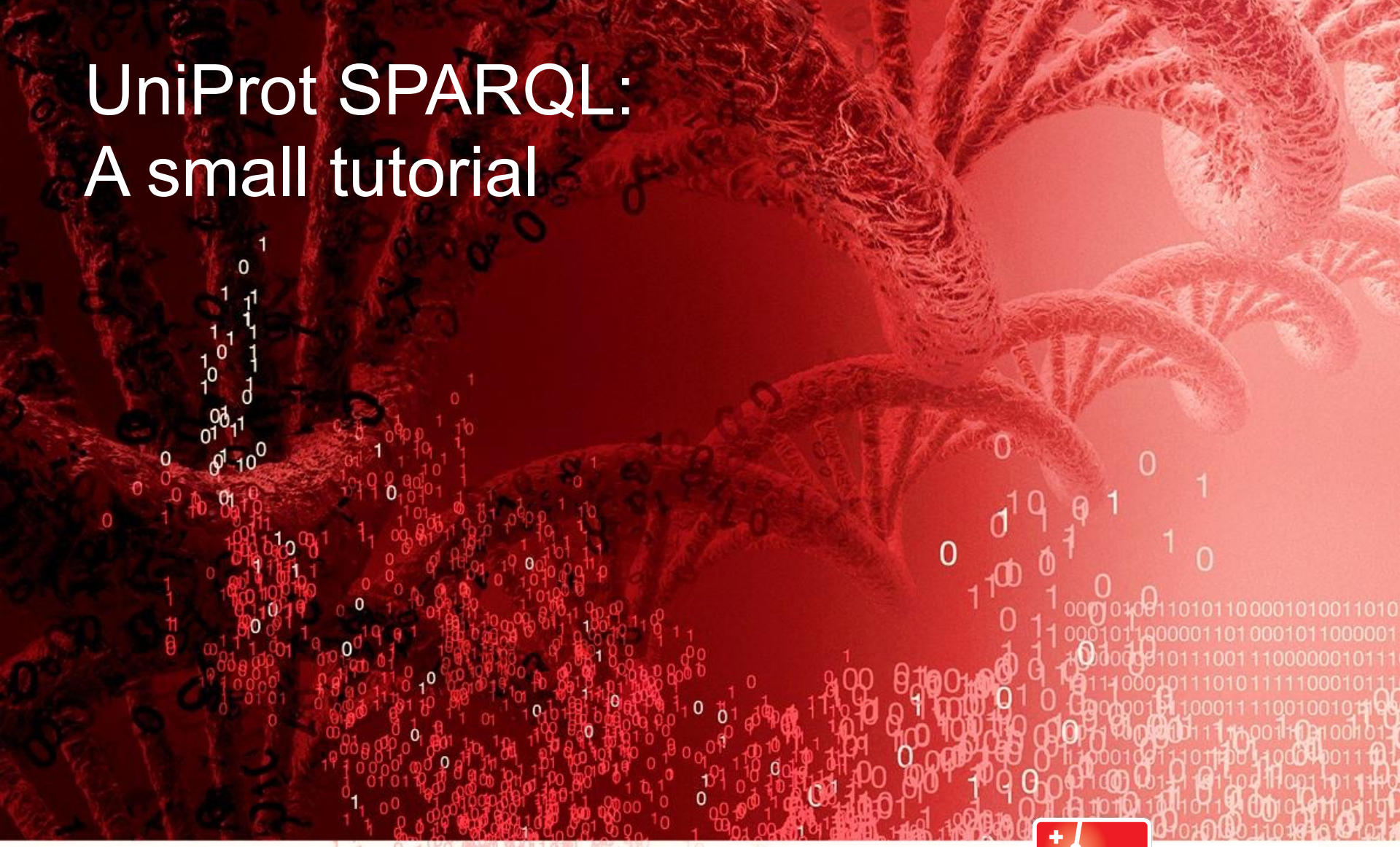
# UniProt SPARQL: A small tutorial



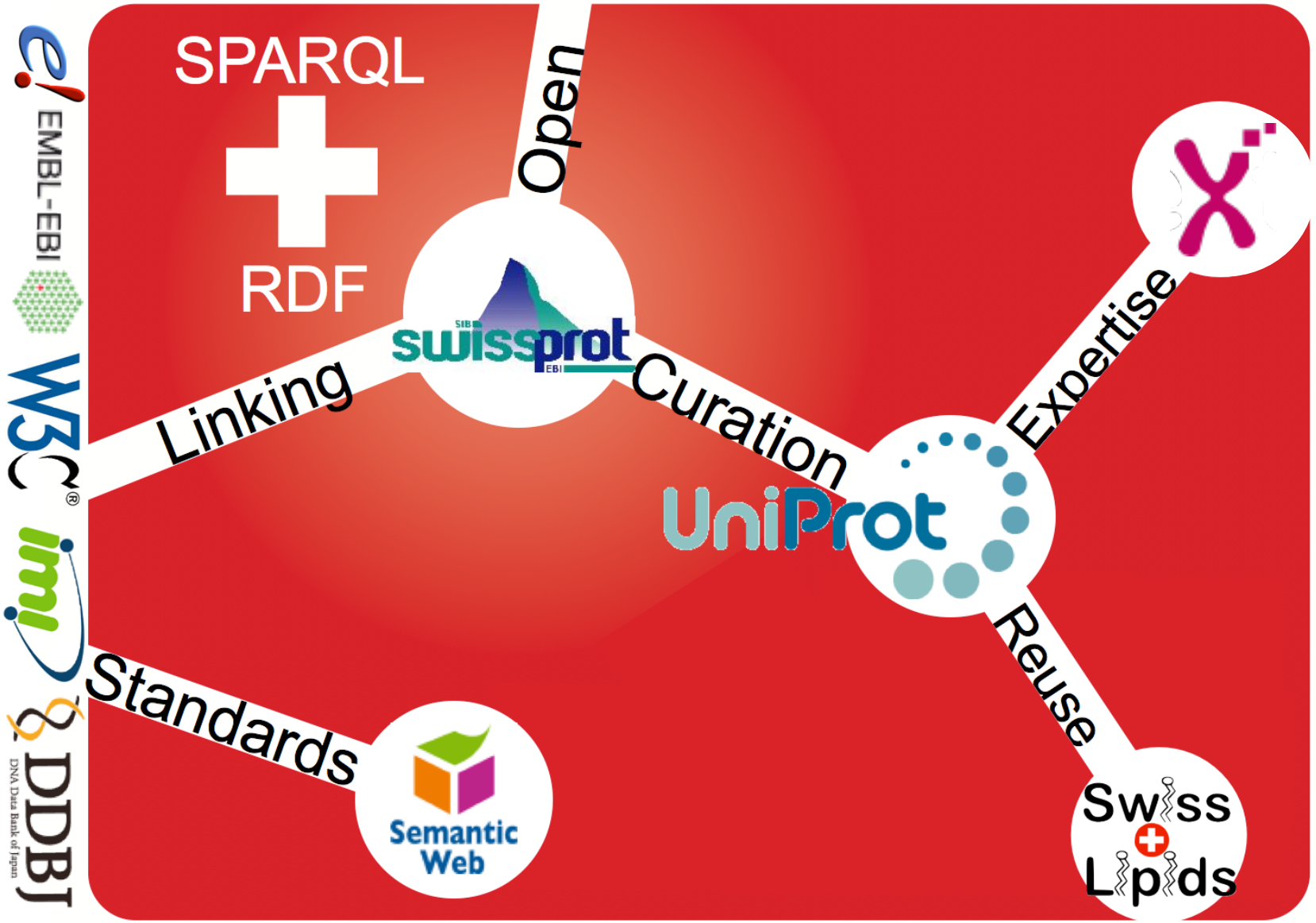
Jerven Bolleman
Swiss-Prot

SIB
Swiss Institute of Bioinformatics

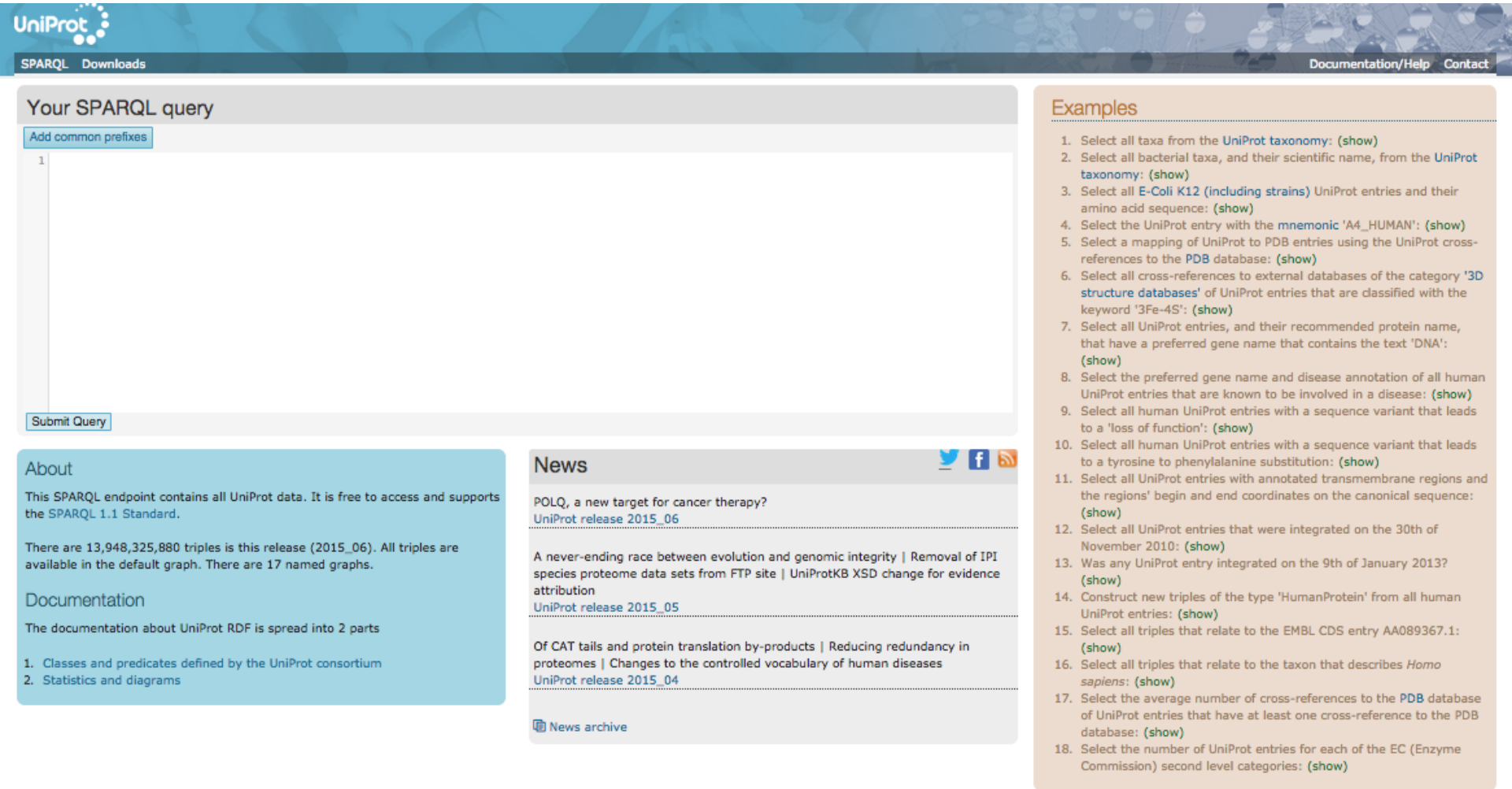
elixir SWITZERLAND

# UniProt SPARQL

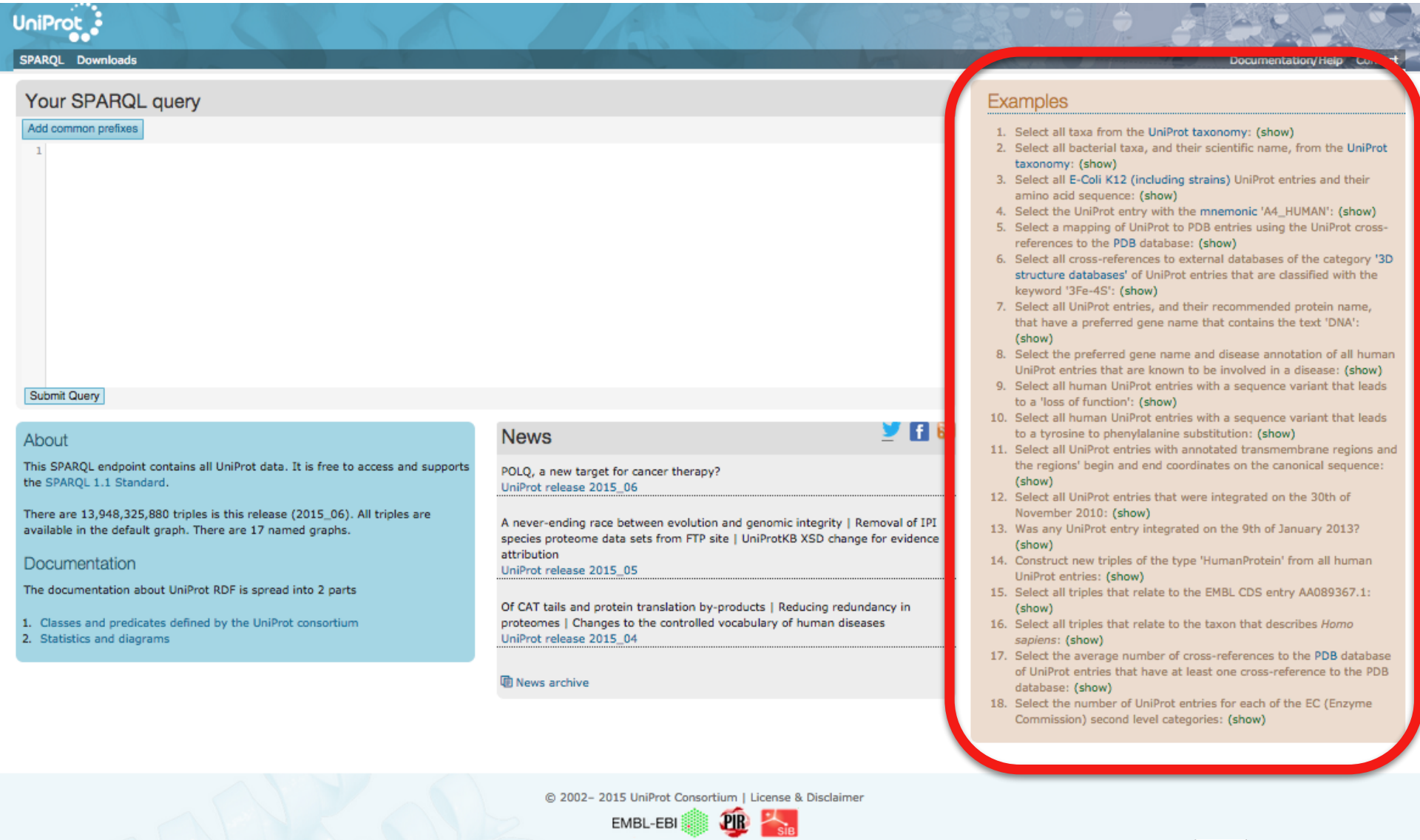
- Next to www.uniprot.org
  - for analytics
  - answering questions
    - not aimed at keyword search
  - your UniProt Data Base in the cloud
- 14 billion triples
  - in 17 graphs
- help@uniprot.org
  - interested in your queries/questions/comments
  - we want you to use our data!
    - all questions are welcome


SPARQL
RDF
Open
Linking
Curation
Expertise
Reuse
Standards
swissprot
UniProt
Semantic Web
Swiss Lipids
EMBL-EBI
W3C
imi
Innovative Medicines Initiative
DDBJ
DNA Data Bank of Japan

# beta.sparql.uniprot.org



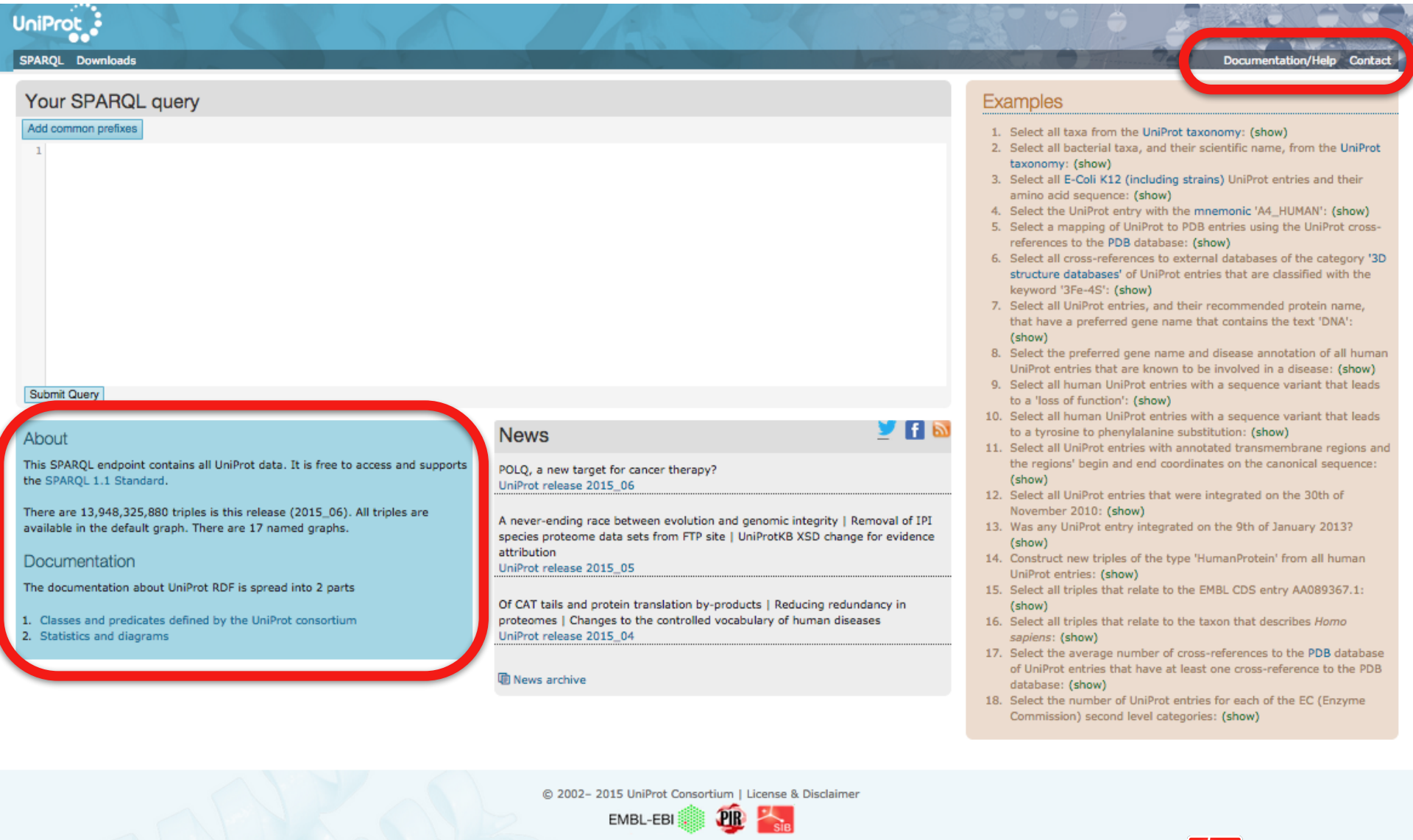
# beta.sparql.uniprot.org






Swiss Institute of Bioinformatics

# beta.sparql.uniprot.org



# http://www.uniprot.org/core/

## Core

### Navigation Panel

| | |
|---|---|
| Classes ( 167 ) | Absorption_Annotation |
| Properties ( 138 ) | |
| Object properties ( 57 ) | alternativeName |
| Datatype properties ( 81 ) | abstract |

### Ontology description

| **http://purl.uniprot.org/core/** (rdf:type owl:Ontology ) | |
|---|---|
| rdfs:comment | Properties and classes used for protein annotation. xsd:string |

### Classes

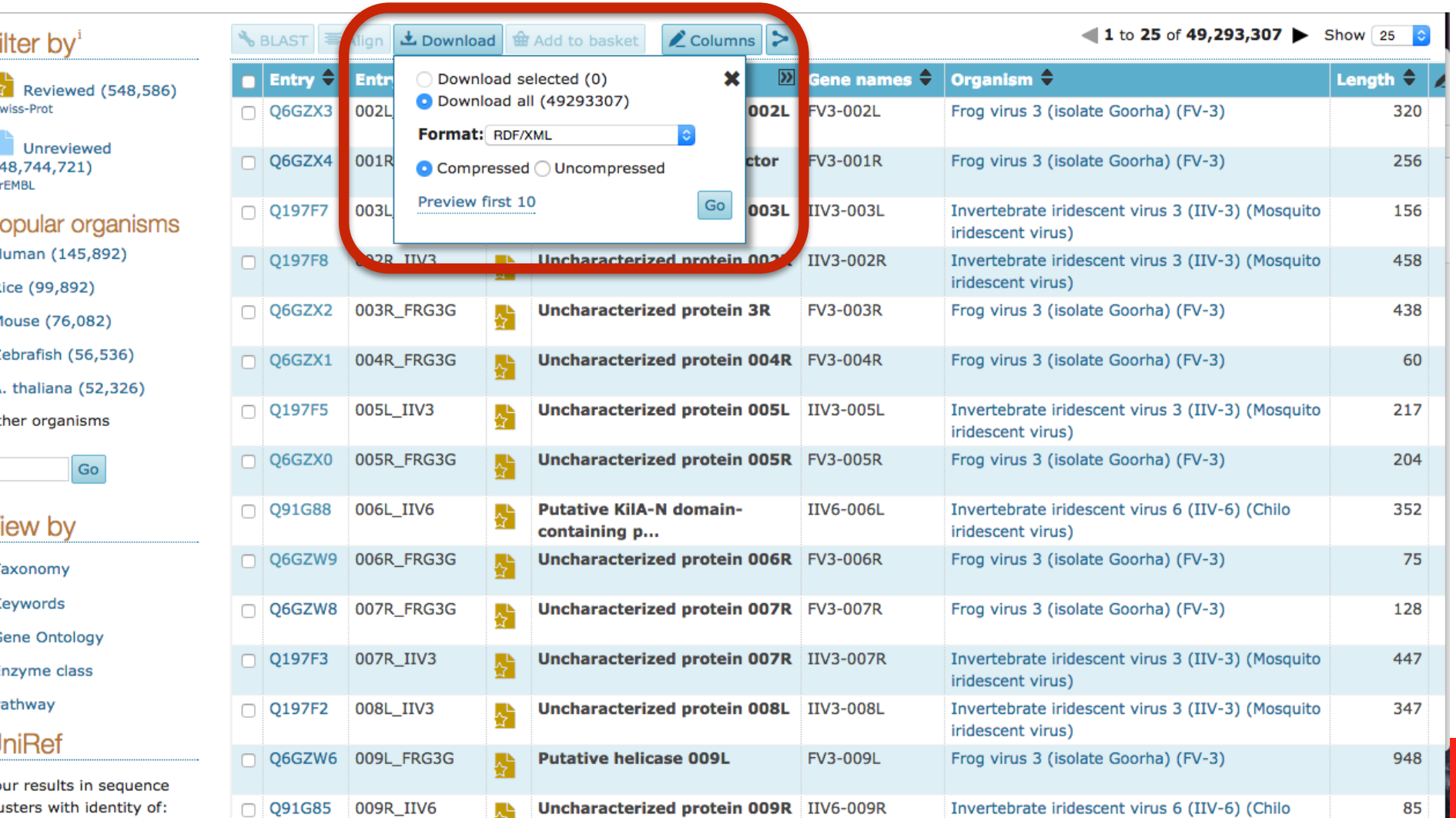
| **Absorption_Annotation** (rdf:type owl:Class ) | |
|---|---|
| rdfs:label | Absorption xsd:string |
| rdfs:subClassOf | Biophysicochemical_Annotation |
| rdfs:comment | Indicates the wavelength in nm at which photoreactive proteins such as opsins and DNA photolyases show maximal absorption. xsd:string |

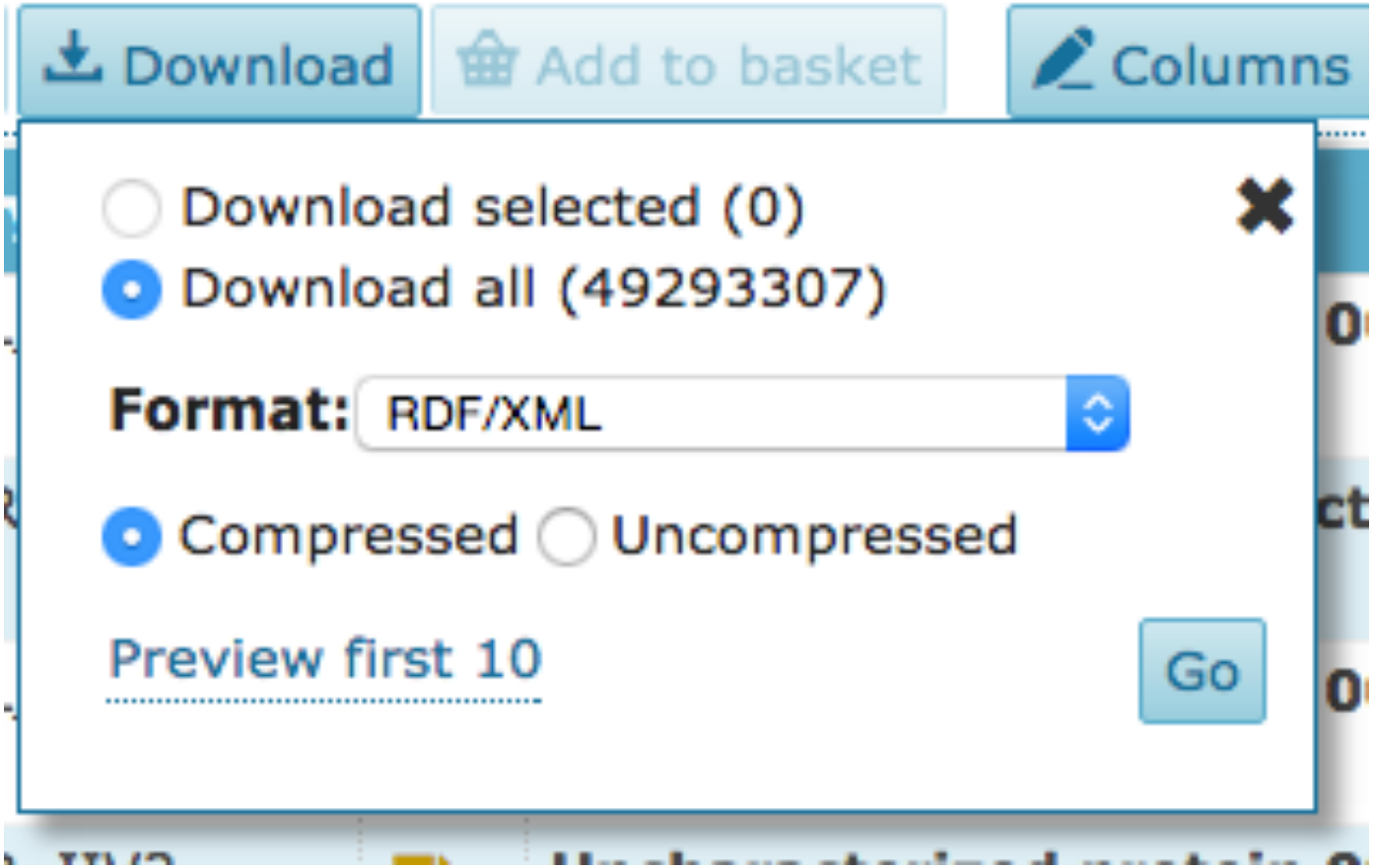
| **Active_Site_Annotation** (rdf:type owl:Class ) | |
|---|---|
| rdfs:label | Active Site xsd:string |
| rdfs:subClassOf | Site_Annotation |
| rdfs:comment | Amino acid(s) involved in the activity of an enzyme. xsd:string |

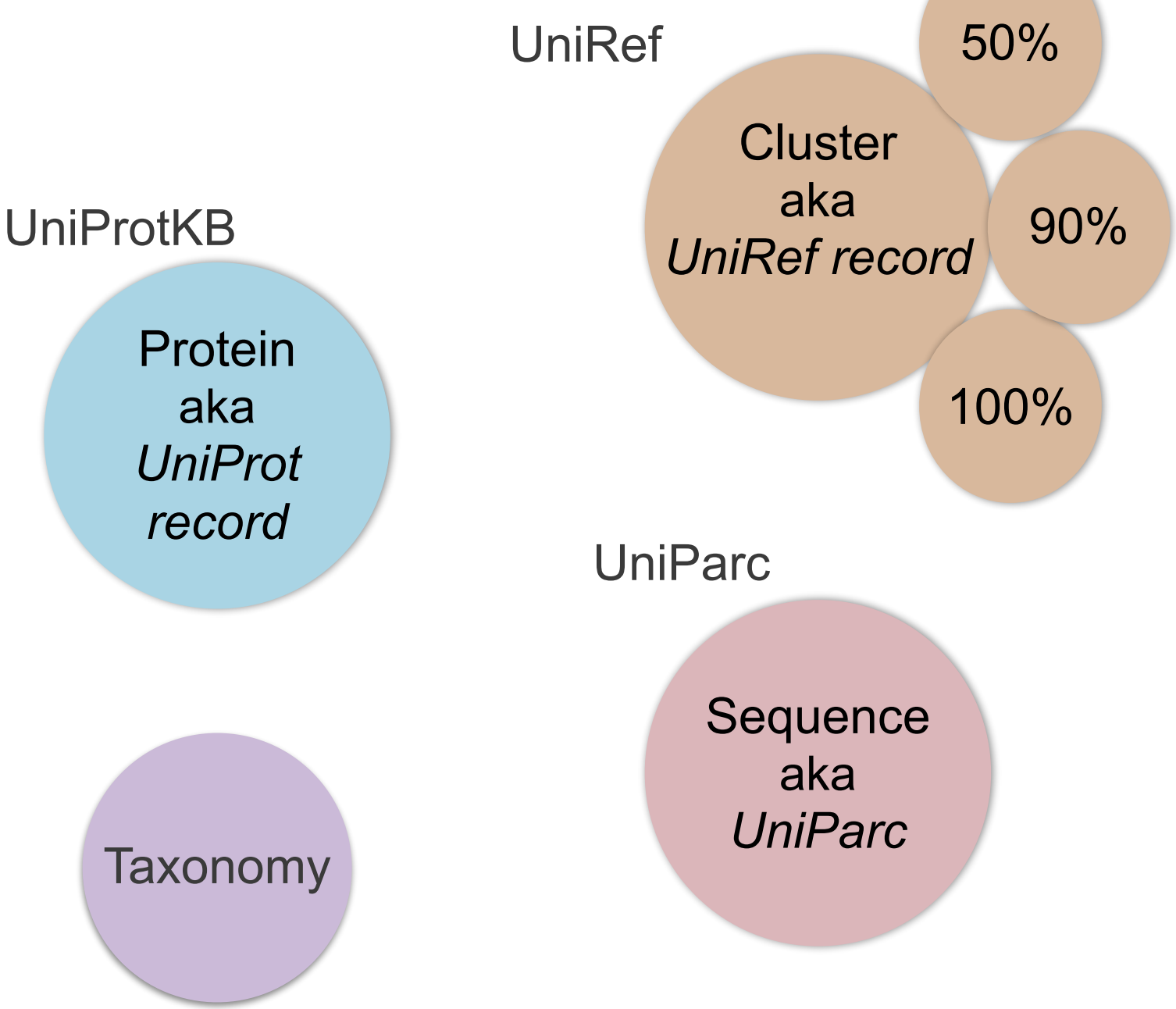
# http://www.uniprot.org/format/
!! work in progress !!

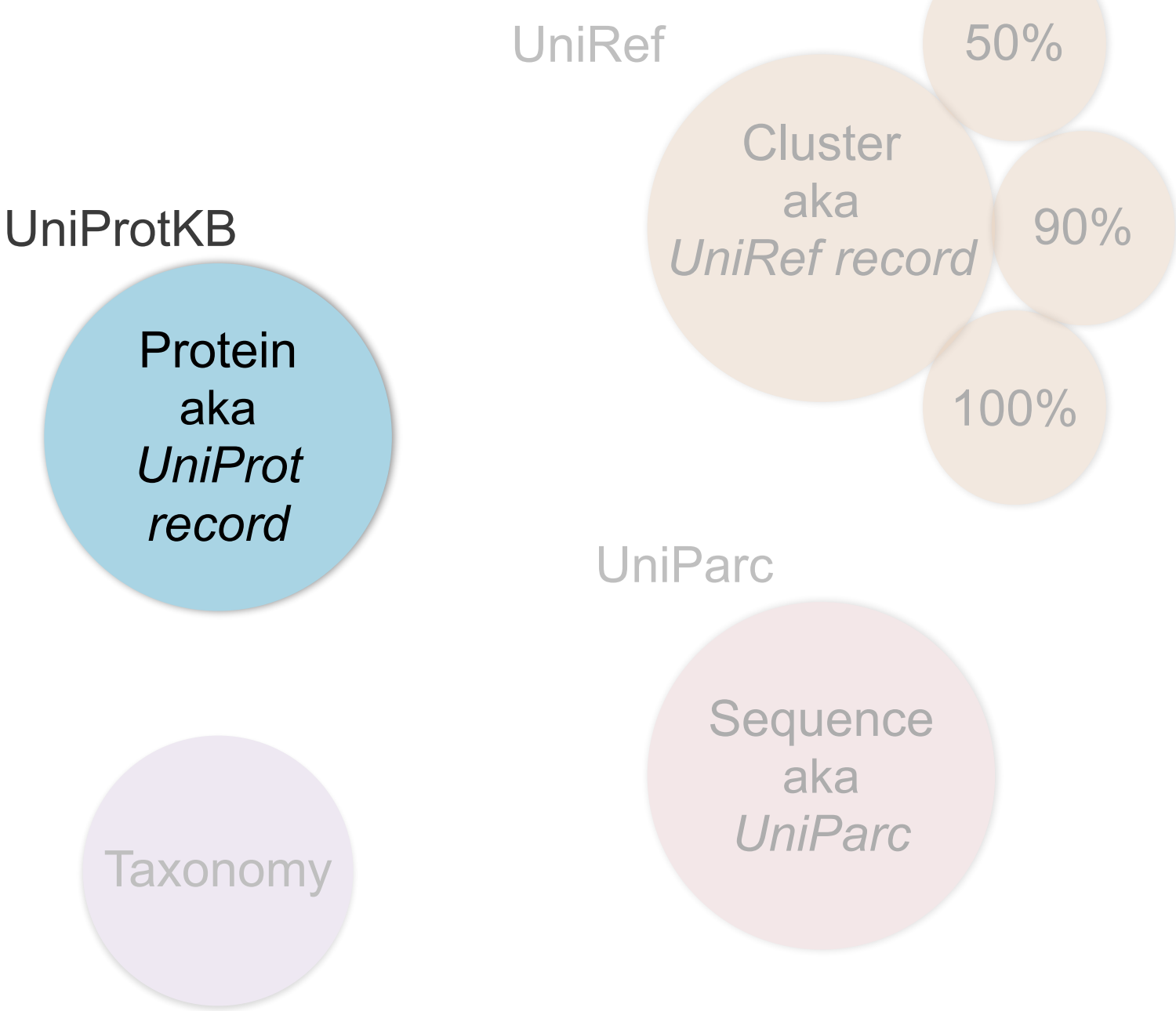
- Description of the different formats
  - Examples in Turtle and JSON-LD
- Bug reports very welcome
- Focuses on UniProtKB

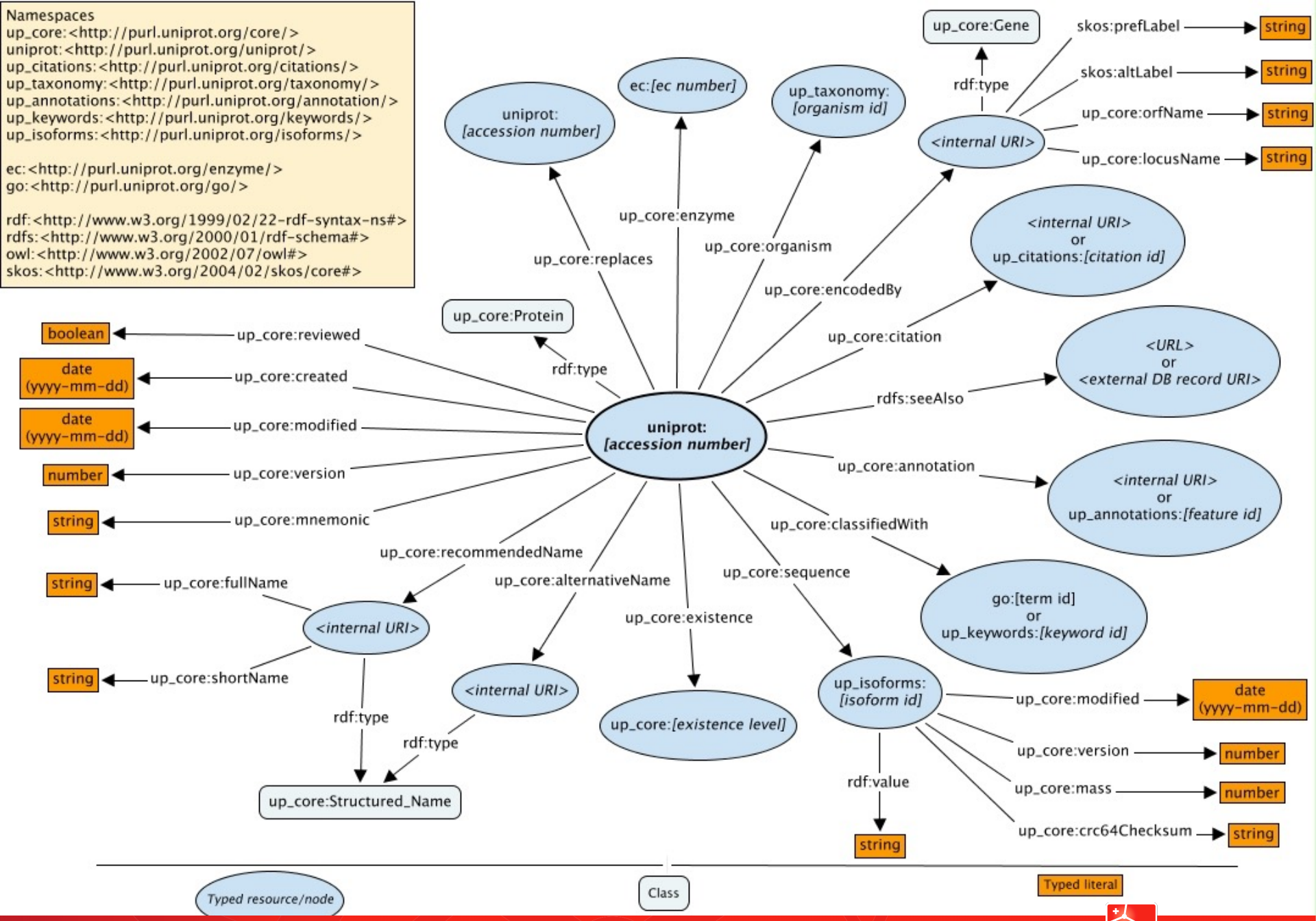
# RDF can be downloaded from uniprot.org

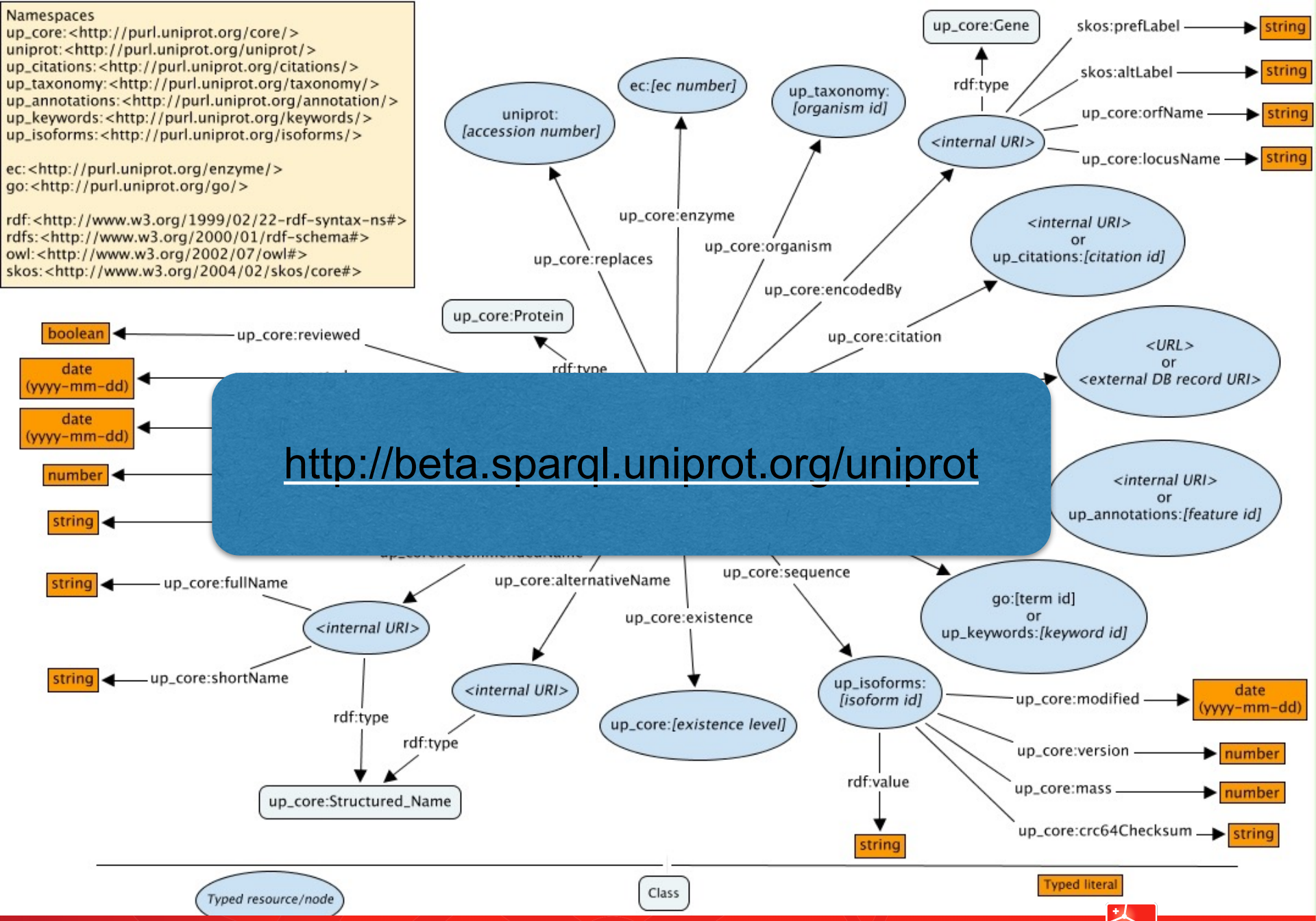



Download
Add to basket
Columns
Download selected (0)
Download all (49293307)
Format:
RDF/XML
Compressed
Uncompressed
Preview first 10
Go

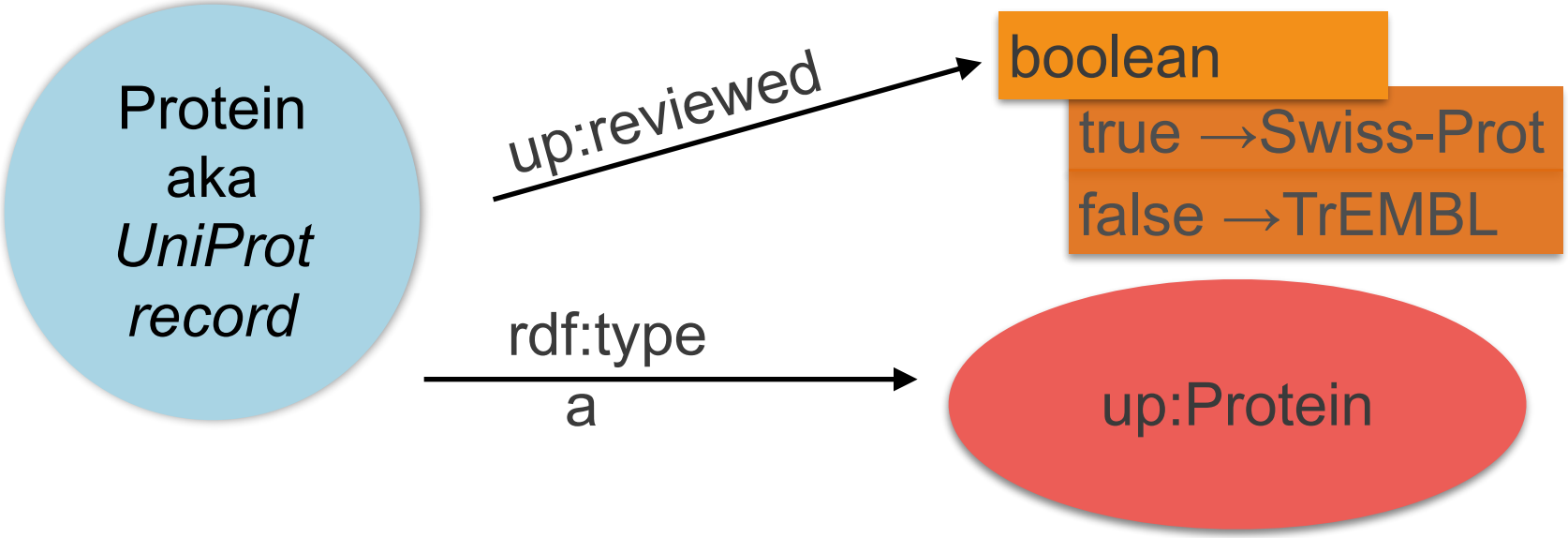

UniRef
50%
Cluster
aka
*UniRef record*
90%
100%
UniProtKB
Protein
aka
*UniProt*
*record*
UniParc
Sequence
aka
*UniParc*
Taxonomy


UniRef
50%
Cluster
aka
*UniRef record*
90%
100%
UniProtKB
Protein
aka
*UniProt*
*record*
UniParc
Sequence
aka
*UniParc*
Taxonomy

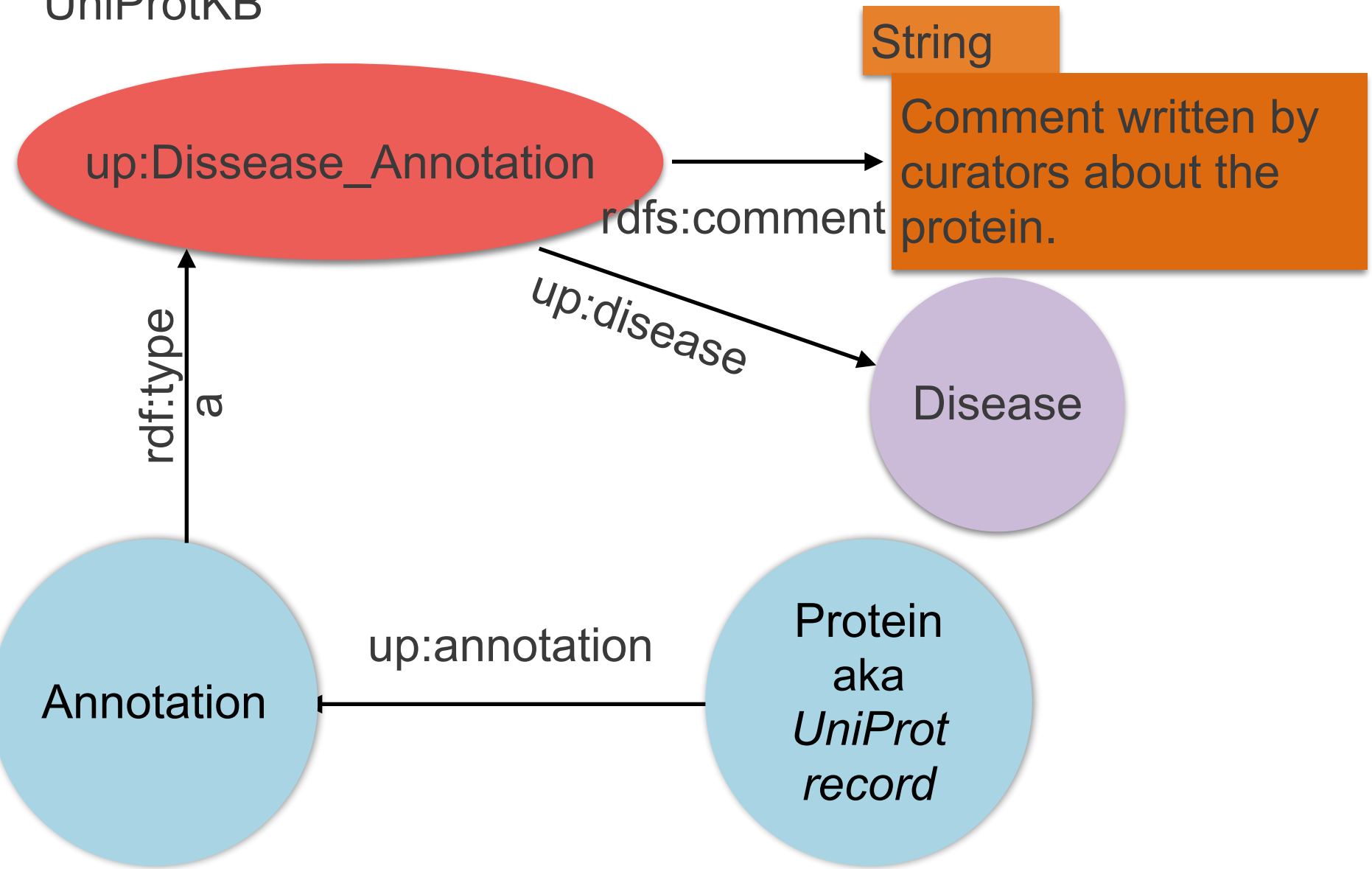

Namespaces
up_core:<http://purl.uniprot.org/core/>
uniprot:<http://purl.uniprot.org/uniprot/>
up_citations:<http://purl.uniprot.org/citations/>
up_taxonomy:<http://purl.uniprot.org/taxonomy/>
up_annotations:<http://purl.uniprot.org/annotation/>
up_keywords:<http://purl.uniprot.org/keywords/>
up_isoforms:<http://purl.uniprot.org/isoforms/>
ec:<http://purl.uniprot.org/enzyme/>
go:<http://purl.uniprot.org/go/>
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#>
skos:<http://www.w3.org/2004/02/skos/core#>
uniprot: [accession number]
ec:[ec number]
up_taxonomy: [organism id]
up_core:Gene
rdf:type
skos:prefLabel
string
skos:altLabel
string
up_core:orfName
string
up_core:locusName
string
<internal URI>
up_core:enzyme
up_core:organism
up_core:replaces
up_core:encodedBy
<internal URI> or up_citations:[citation id]
up_core:citation
up_core:Protein
rdf:type
boolean
up_core:reviewed
date (yyyy-mm-dd)
up_core:created
date (yyyy-mm-dd)
up_core:modified
number
up_core:version
string
up_core:mnemonic
<URL> or <external DB record URI>
rdfs:seeAlso
uniprot: [accession number]
up_core:annotation
<internal URI> or up_annotations:[feature id]
up_core:classifiedWith
go:[term id] or up_keywords:[keyword id]
up_core:recommendedName
up_core:alternativeName
up_core:sequence
up_core:existence
string
up_core:fullName
<internal URI>
string
up_core:shortName
rdf:type
<internal URI>
rdf:type
up_core:Structured_Name
up_core:[existence level]
up_isoforms: [isoform id]
up_core:modified
date (yyyy-mm-dd)
up_core:version
number
up_core:mass
number
up_core:crc64Checksum
string
rdf:value
string
Typed resource/node
Class
Typed literal


Namespaces
up_core:<http://purl.uniprot.org/core/>
uniprot:<http://purl.uniprot.org/uniprot/>
up_citations:<http://purl.uniprot.org/citations/>
up_taxonomy:<http://purl.uniprot.org/taxonomy/>
up_annotations:<http://purl.uniprot.org/annotation/>
up_keywords:<http://purl.uniprot.org/keywords/>
up_isoforms:<http://purl.uniprot.org/isoforms/>
ec:<http://purl.uniprot.org/enzyme/>
go:<http://purl.uniprot.org/go/>
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#>
skos:<http://www.w3.org/2004/02/skos/core#>
http://beta.sparql.uniprot.org/uniprot
up_core:Gene
skos:prefLabel
skos:altLabel
up_core:orfName
up_core:locusName
rdf:type
uniprot: [accession number]
ec:[ec number]
up_taxonomy: [organism id]
<internal URI>
up_core:enzyme
up_core:organism
up_core:replaces
up_core:encodedBy
<internal URI> or up_citations:[citation id]
up_core:Protein
up_core:reviewed
up_core:citation
<URL> or <external DB record URI>
boolean
date (yyyy-mm-dd)
number
string
<internal URI> or up_annotations:[feature id]
up_core:fullName
up_core:alternativeName
up_core:sequence
go:[term id] or up_keywords:[keyword id]
up_core:shortName
up_core:existence
up_core:[existence level]
up_isoforms: [isoform id]
up_core:modified
up_core:version
up_core:mass
up_core:crc64Checksum
rdf:value
up_core:Structured_Name
Typed resource/node
Class
Typed literal

# UniProtKB

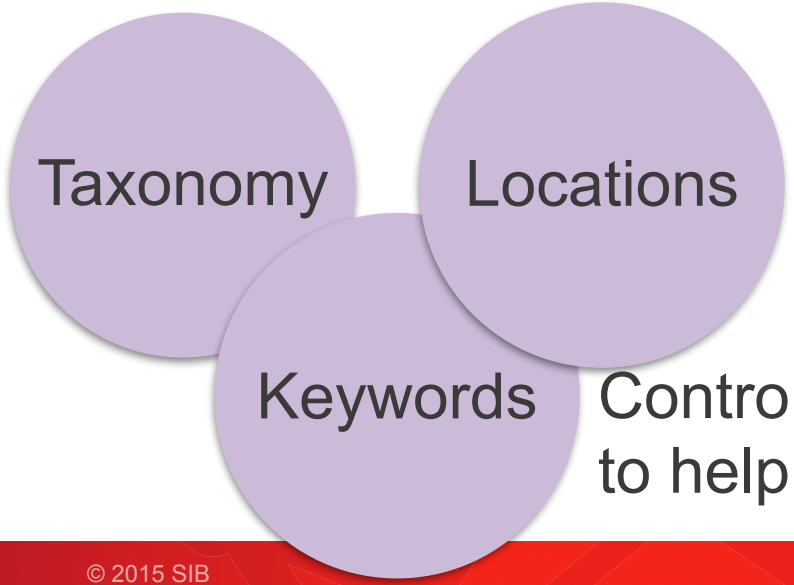


UniProtKB



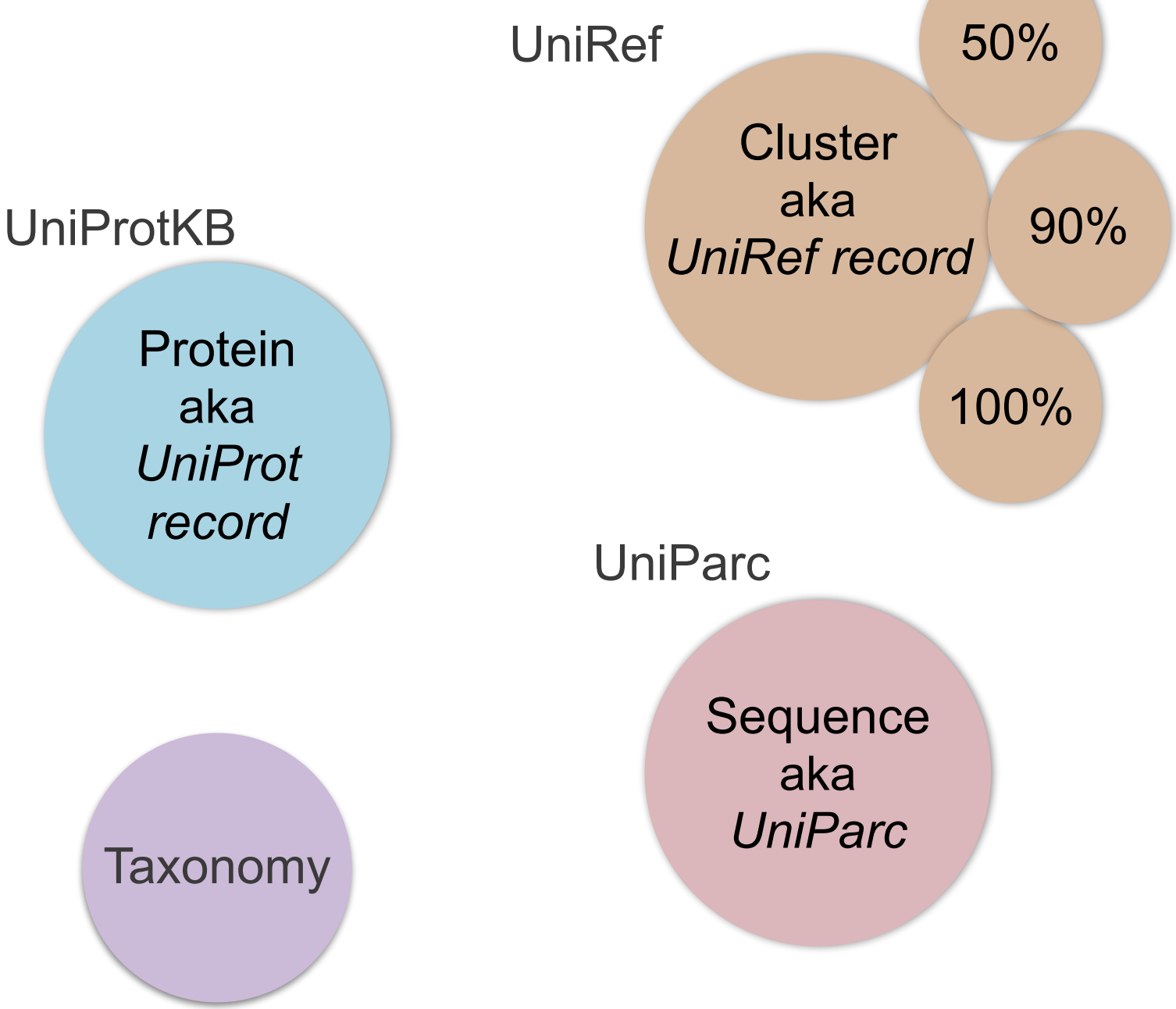
UniProtKB

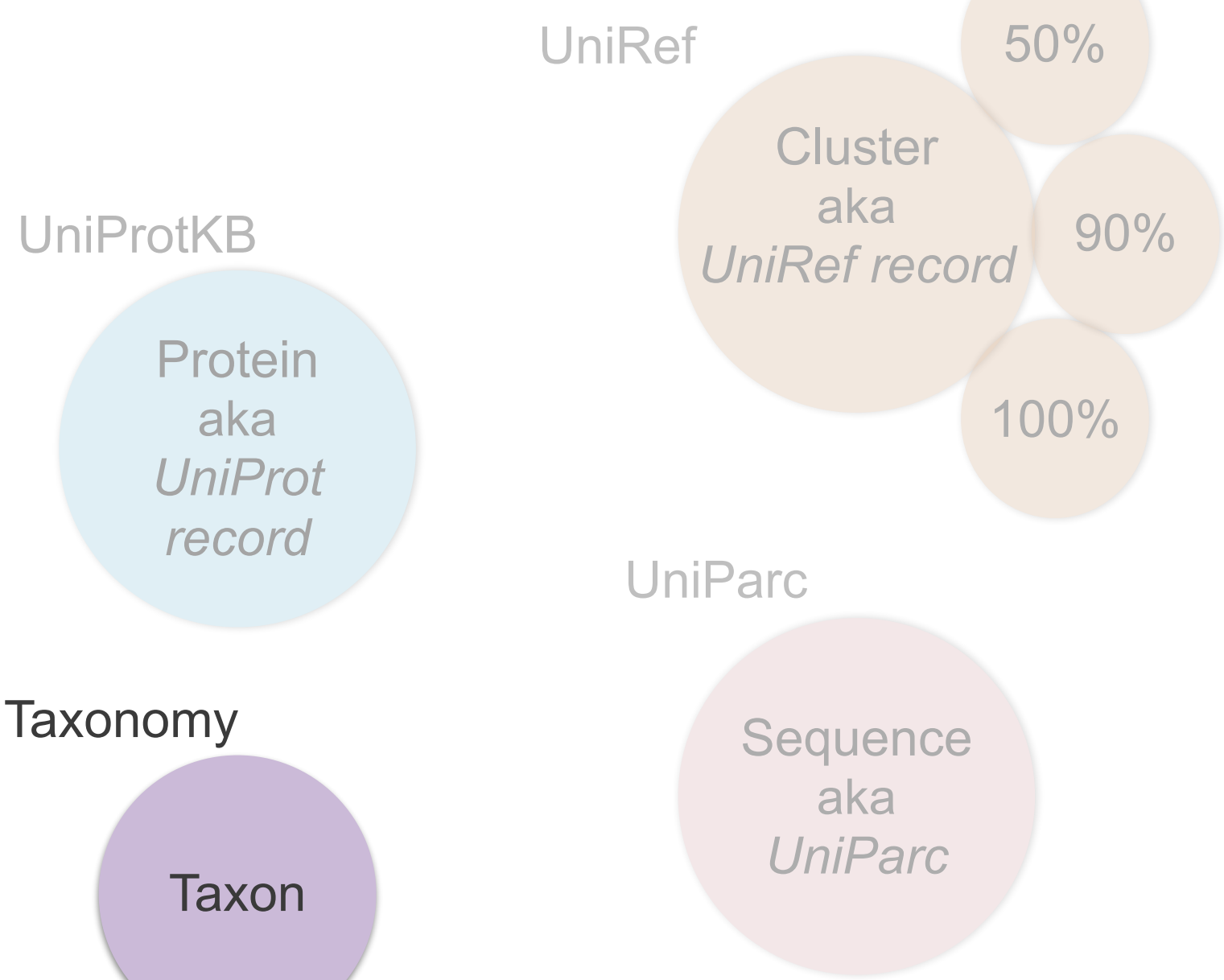
Protein
aka
*UniProt*
*record*

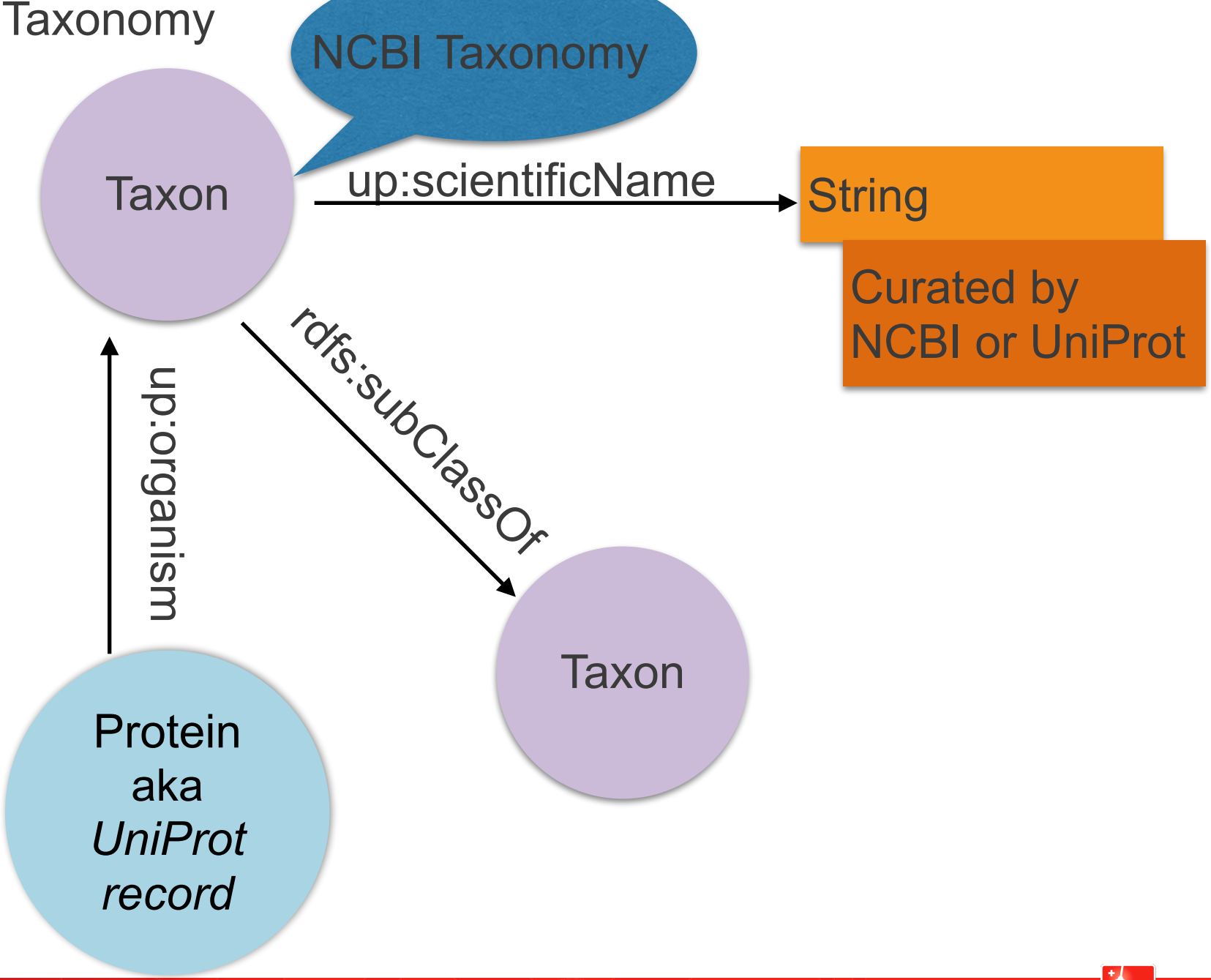


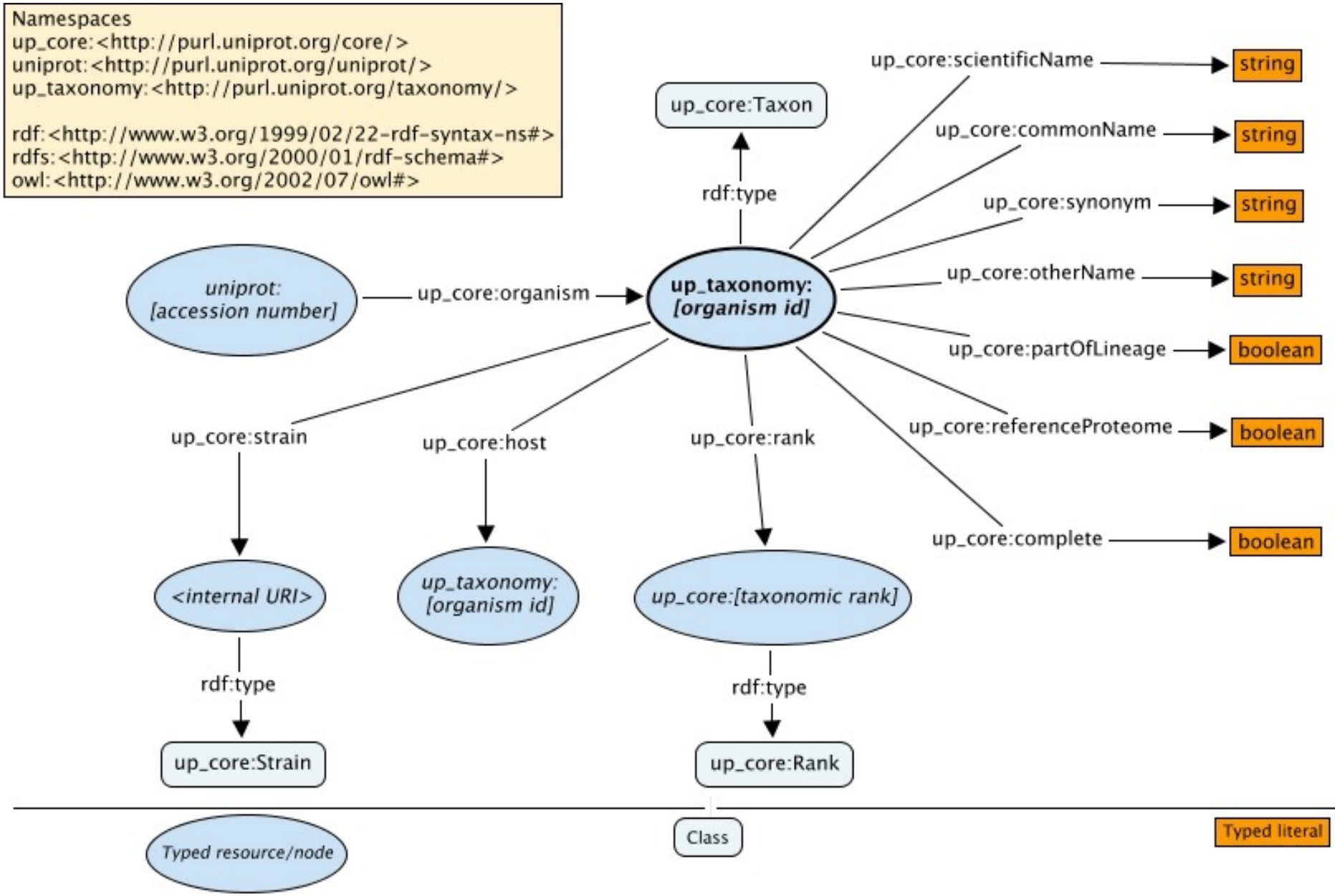

Controlled vocabularies
to help you find information

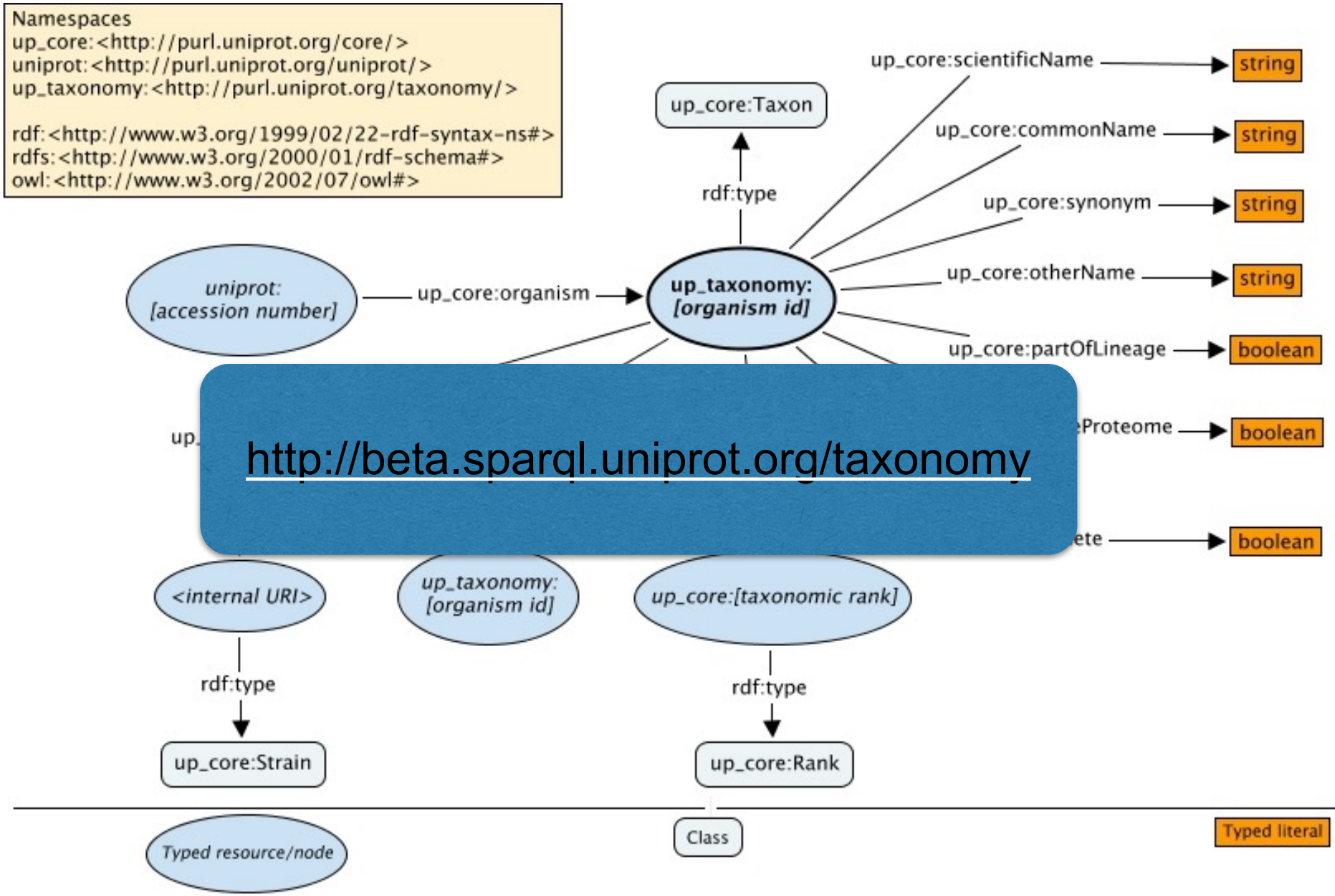

Reusing ontologies
for ease of integration

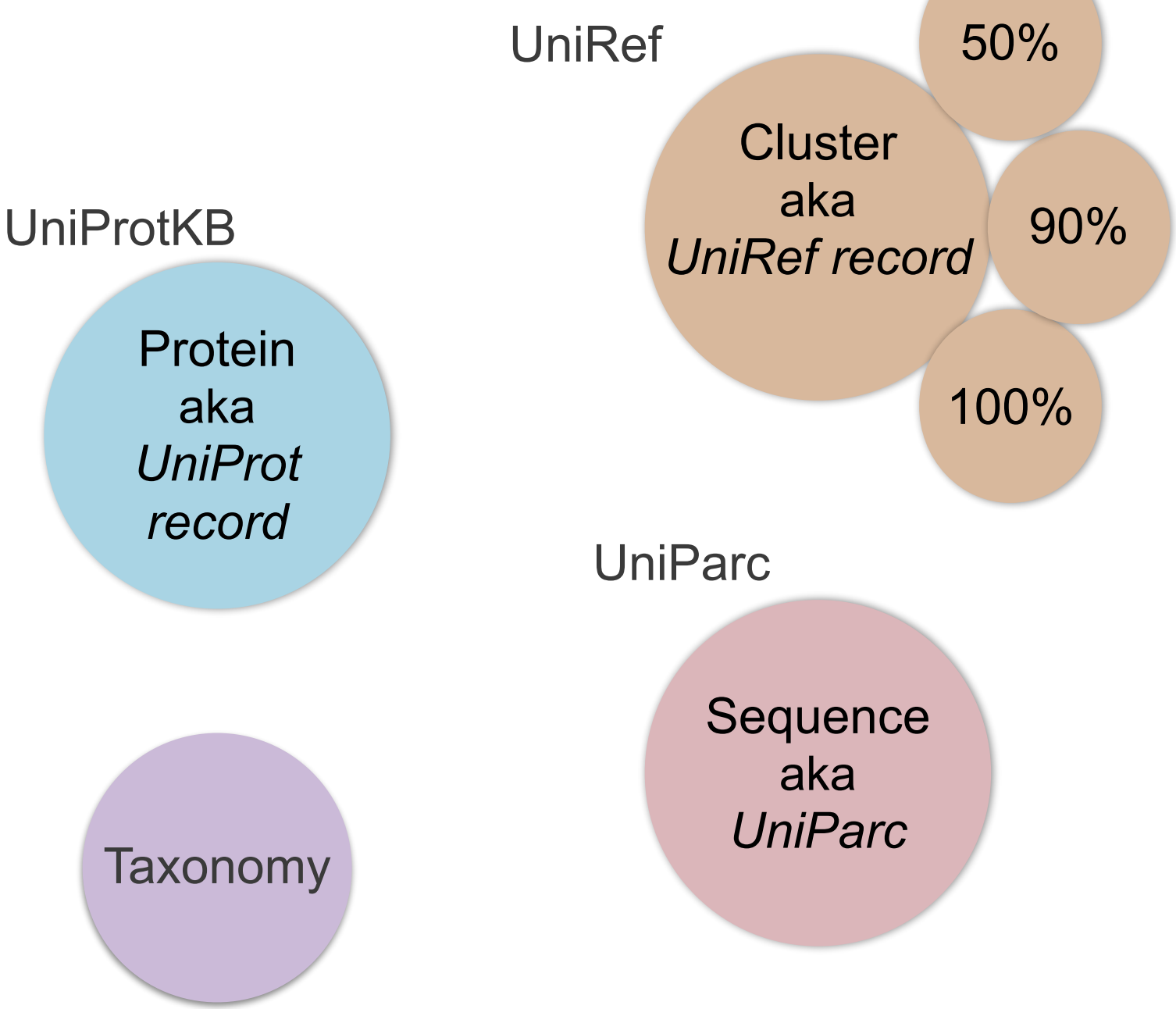

UniRef
50%
Cluster
aka
*UniRef record*
90%
100%
UniProtKB
Protein
aka
*UniProt*
*record*
UniParc
Sequence
aka
*UniParc*
Taxonomy

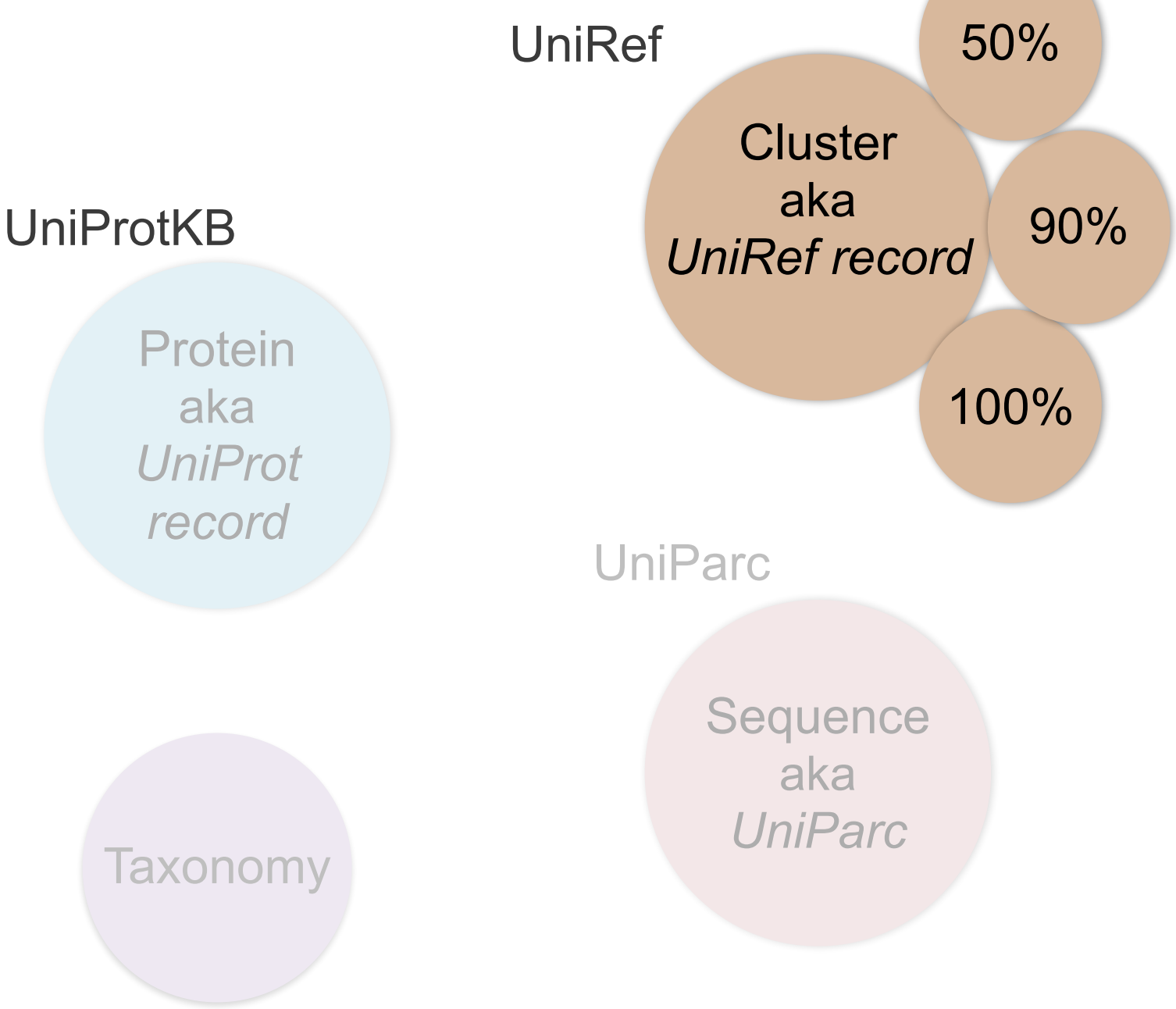

UniRef
50%
Cluster
aka
*UniRef record*
90%
100%
UniProtKB
Protein
aka
*UniProt*
*record*
UniParc
Sequence
aka
*UniParc*
Taxonomy
Taxon

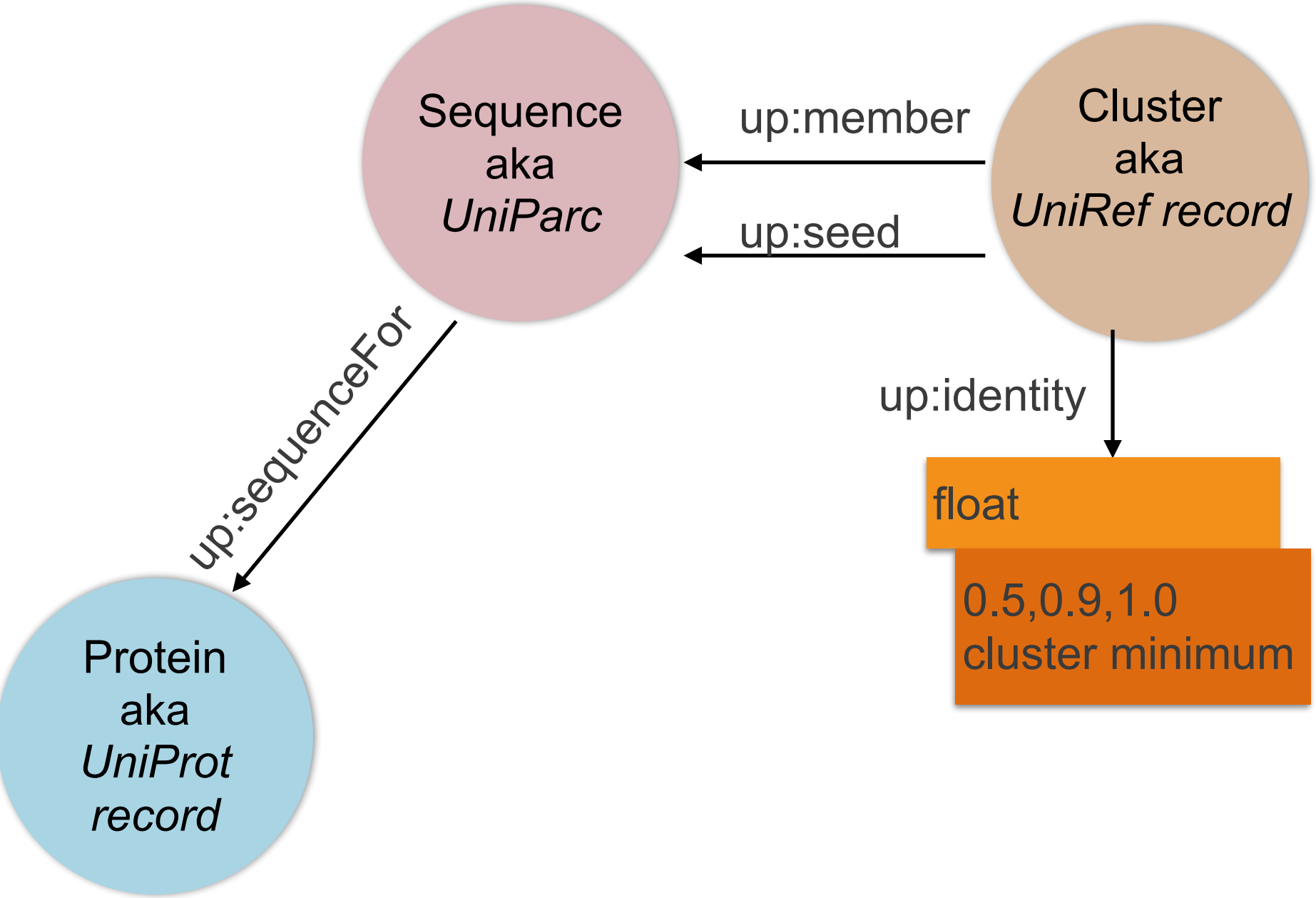

Taxonomy
NCBI Taxonomy
Taxon
up:scientificName
String
Curated by
NCBI or UniProt
rdfs:subClassOf
up:organism
Taxon
Protein
aka
UniProt
record

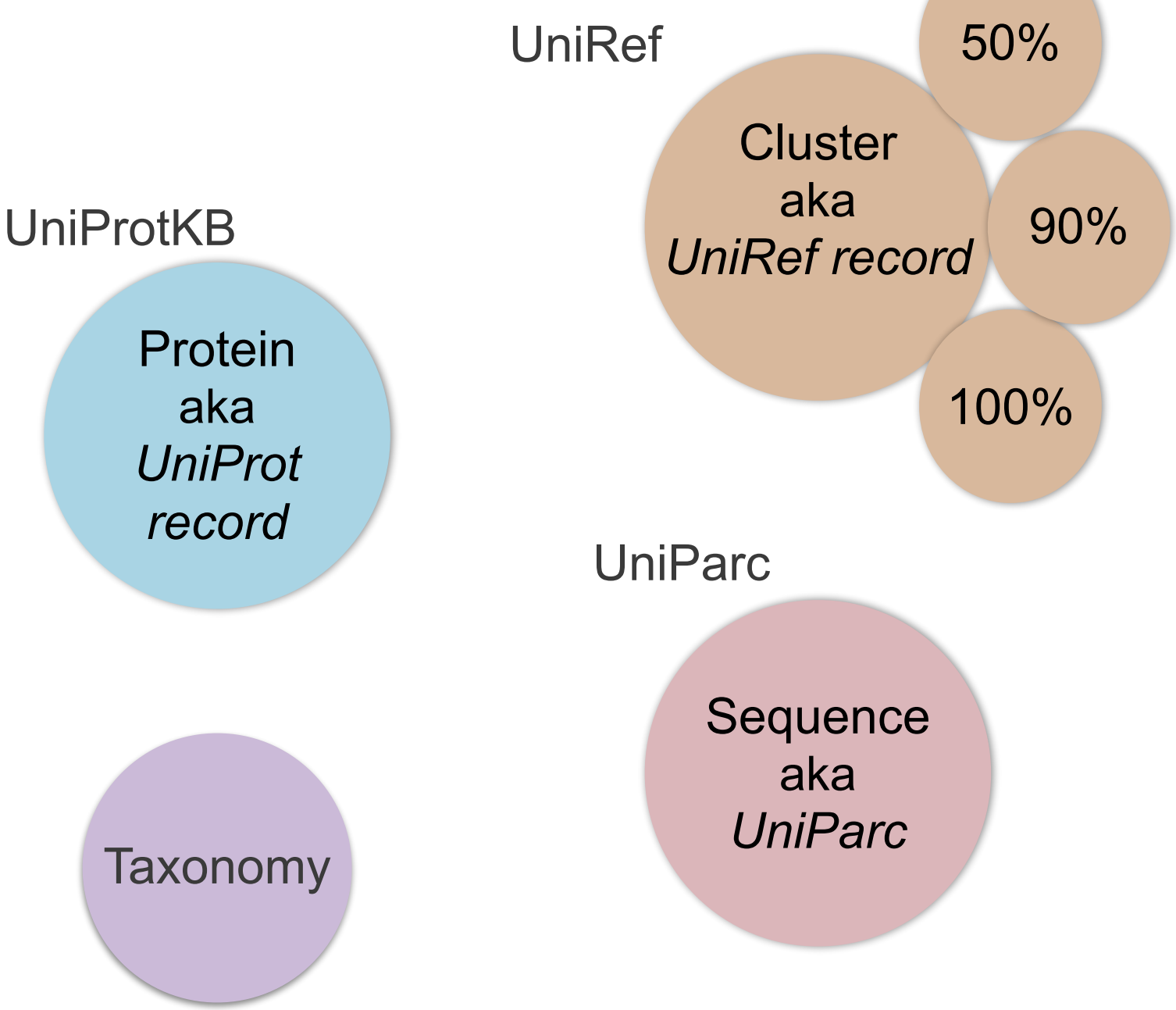

Namespaces
up_core:<http://purl.uniprot.org/core/>
uniprot:<http://purl.uniprot.org/uniprot/>
up_taxonomy:<http://purl.uniprot.org/taxonomy/>
rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
rdfs:<http://www.w3.org/2000/01/rdf-schema#>
owl:<http://www.w3.org/2002/07/owl#>
up_core:Taxon
rdf:type
up_core:scientificName
string
up_core:commonName
string
up_core:synonym
string
up_core:otherName
string
uniprot:
[accession number]
up_core:organism
up_taxonomy:
[organism id]
up_core:partOfLineage
boolean
up_core:referenceProteome
boolean
up_core:complete
boolean
up_core:strain
up_core:host
up_core:rank
<internal URI>
up_taxonomy:
[organism id]
up_core:[taxonomic rank]
rdf:type
rdf:type
up_core:Strain
up_core:Rank
Typed resource/node
Class
Typed literal

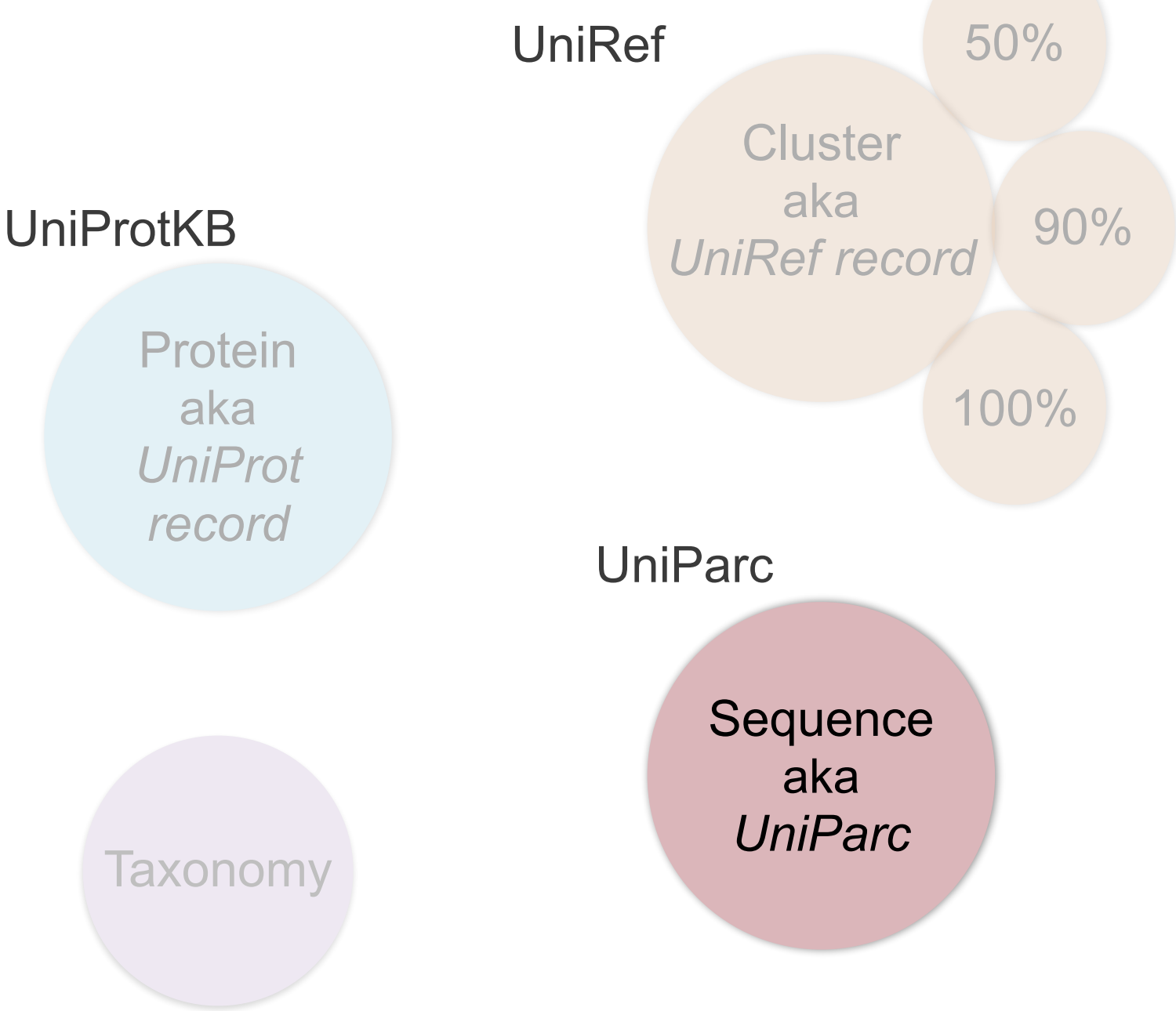

Namespaces
up_core:<http://purl.uniprot.org/core/>
uniprot:<http://purl.uniprot.org/uniprot/>
up_taxonomy:<http://purl.uniprot.org/taxonomy/>
rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
rdfs:<http://www.w3.org/2000/01/rdf-schema#>
owl:<http://www.w3.org/2002/07/owl#>
up_core:Taxon
rdf:type
up_core:scientificName
string
up_core:commonName
string
up_core:synonym
string
up_core:otherName
string
uniprot:
[accession number]
up_core:organism
up_taxonomy:
[organism id]
up_core:partOfLineage
boolean
Proteome
boolean
ete
boolean
http://beta.sparql.uniprot.org/taxonomy
<internal URI>
up_taxonomy:
[organism id]
up_core:[taxonomic rank]
rdf:type
rdf:type
up_core:Strain
up_core:Rank
Typed resource/node
Class
Typed literal

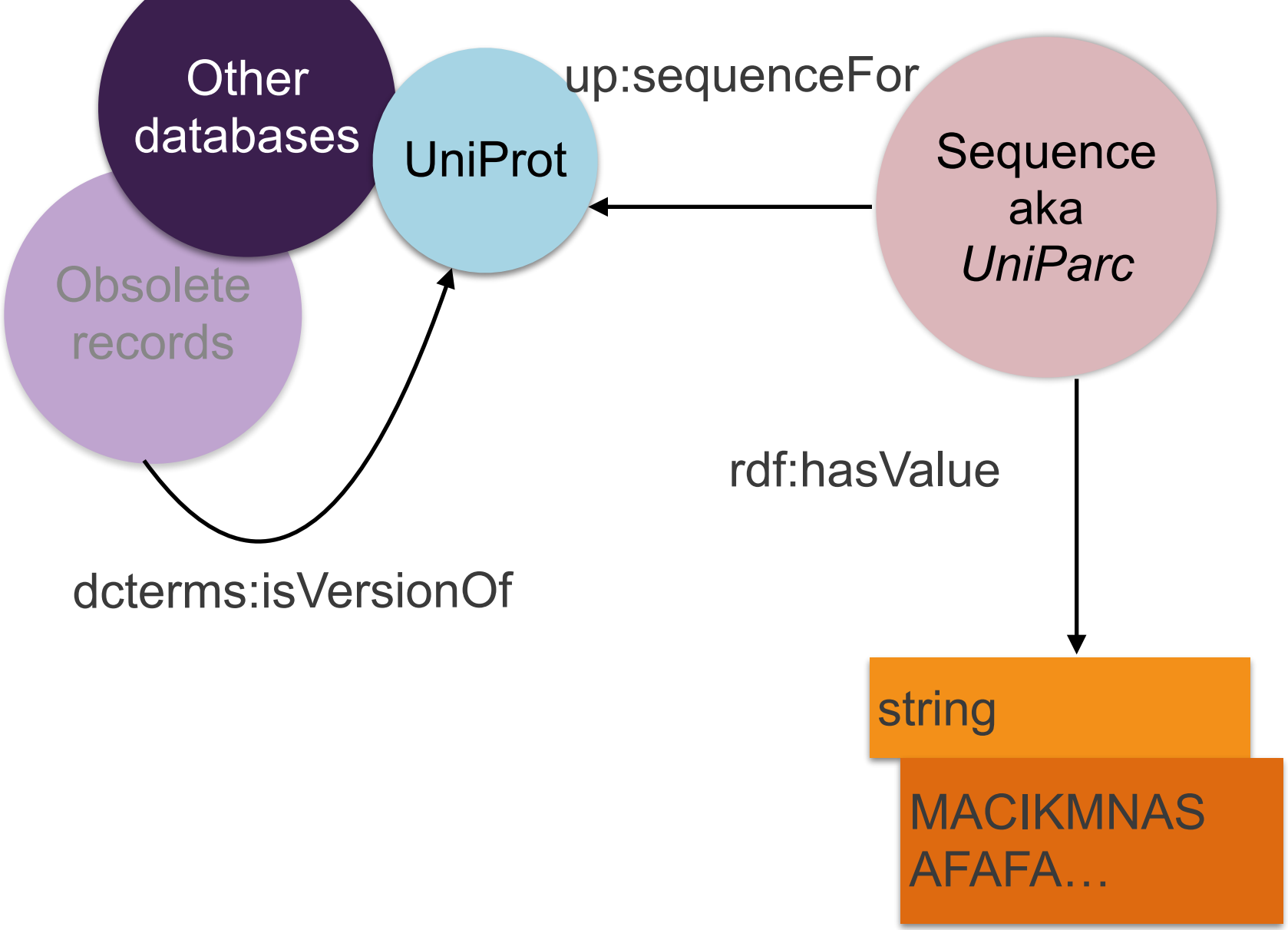

UniRef
50%
Cluster
aka
*UniRef record*
90%
100%
UniProtKB
Protein
aka
*UniProt*
*record*
UniParc
Sequence
aka
*UniParc*
Taxonomy


UniRef
50%
Cluster
aka
*UniRef record*
90%
100%
UniProtKB
Protein
aka
*UniProt*
*record*
UniParc
Sequence
aka
*UniParc*
Taxonomy


Sequence
aka
UniParc
up:member
up:seed
Cluster
aka
UniRef record
up:sequenceFor
up:identity
float
0.5,0.9,1.0
cluster minimum
Protein
aka
UniProt
record


UniRef
50%
Cluster
aka
*UniRef record*
90%
100%
UniProtKB
Protein
aka
*UniProt*
*record*
UniParc
Sequence
aka
*UniParc*
Taxonomy


UniRef
50%
Cluster
aka
*UniRef record*
90%
100%
UniProtKB
Protein
aka
*UniProt*
*record*
UniParc
Sequence
aka
*UniParc*
Taxonomy


Other databases
UniProt
Obsolete records
up:sequenceFor
Sequence aka UniParc
rdf:hasValue
dcterms:isVersionOf
string
MACIKMNAS AFAFA…

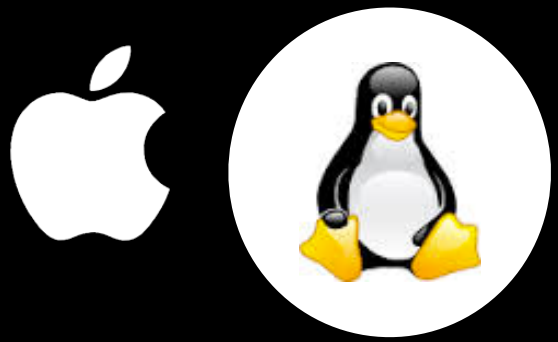
# We are using EPO and EPO receptors as training data

- http://www.uniprot.org/uniprot/P19235
  - EPO_HUMAN
- http://www.uniprot.org/uniprot/P01588
  - EPOR_HUMAN
- http://www.uniprot.org/uniparc/UPI000012A0AD
- http://www.uniprot.org/uniref/UniRef100_P19235
- etc…

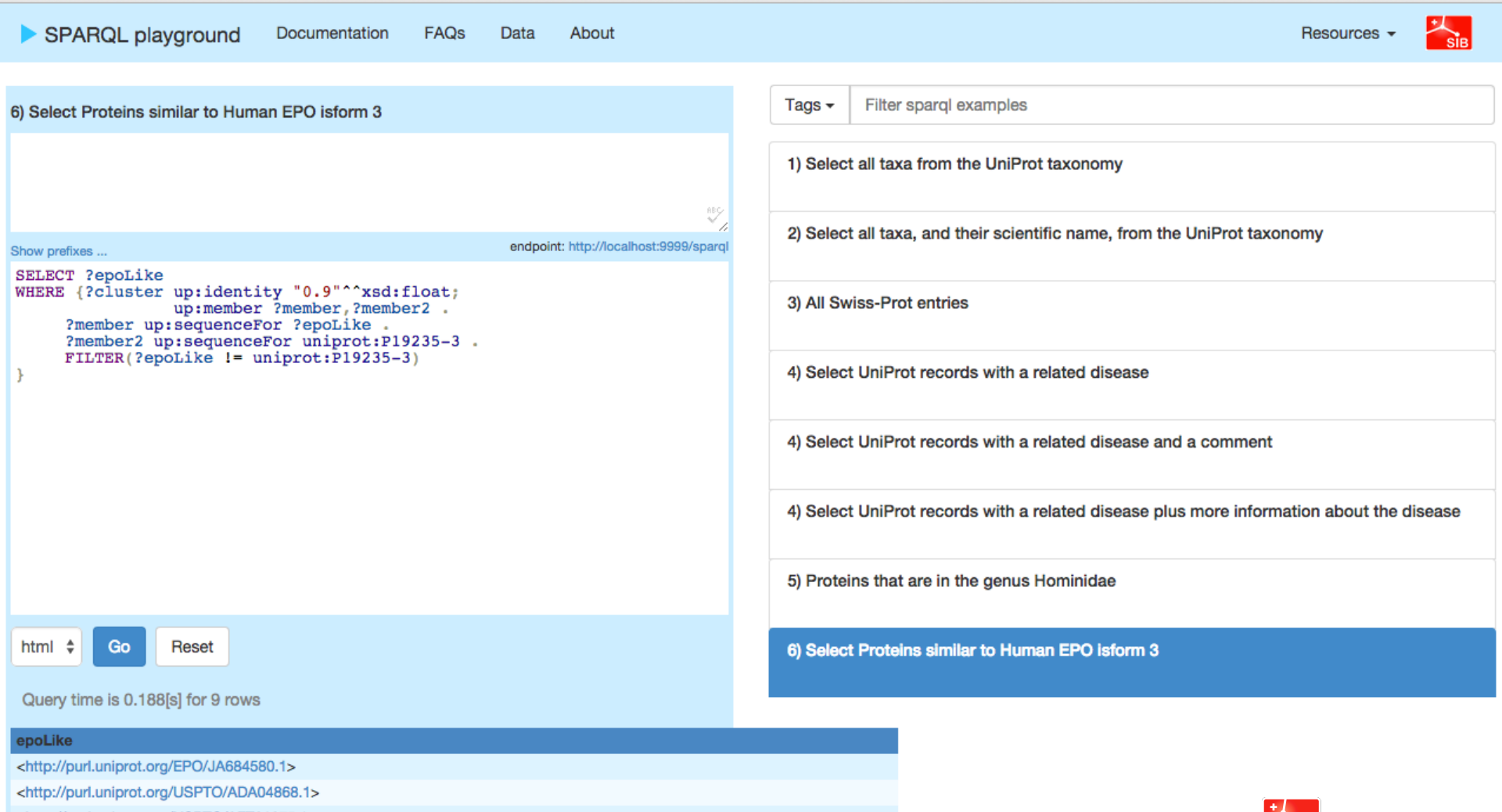
```
./start-uniprot.sh
```

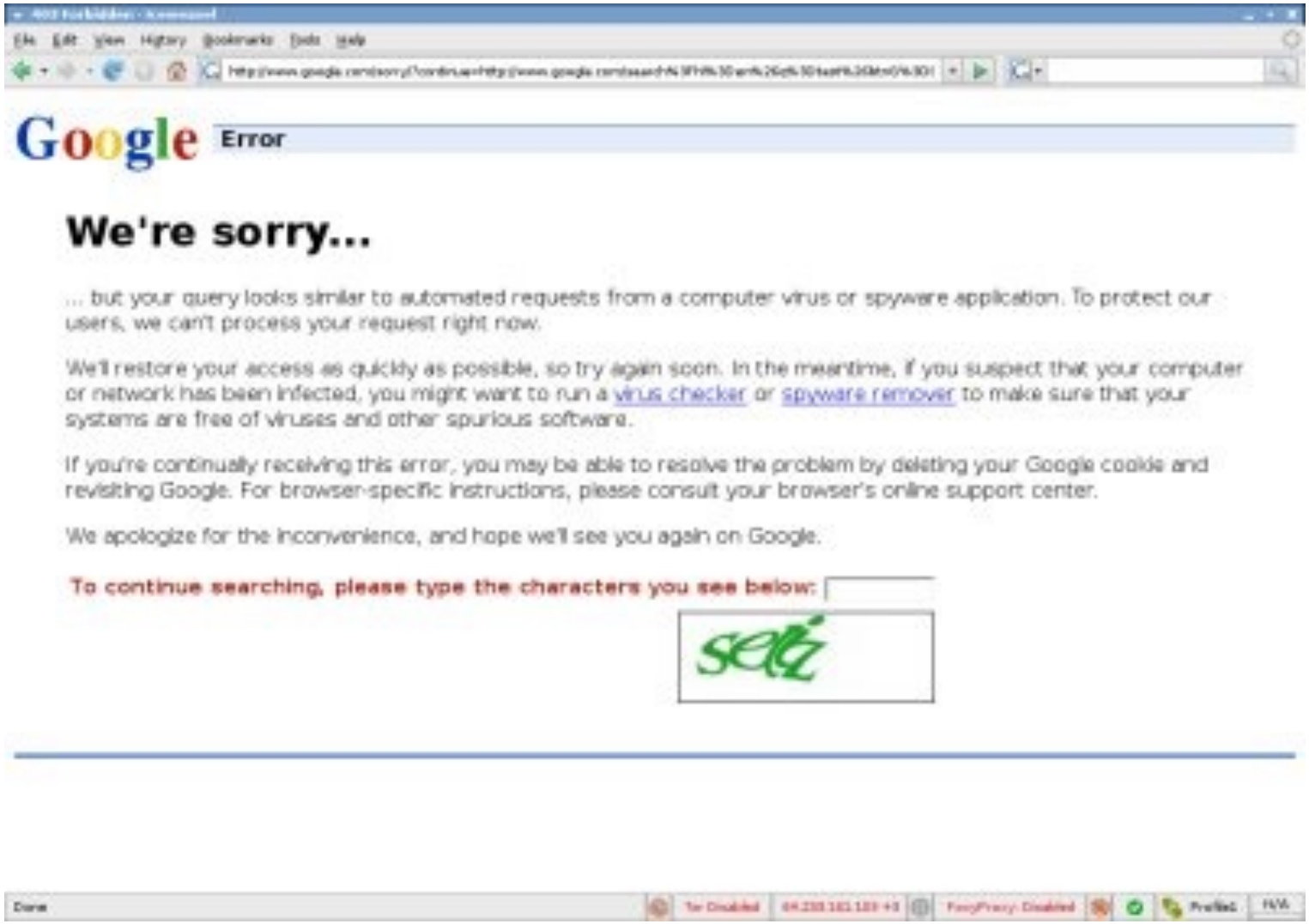
or

```
./start-uniprot.bat
```



http://localhost:9999
http://bc2-2015.sparql.uniprot.org/




Google
Error
We're sorry...
... but your query looks similar to automated requests from a computer virus or spyware application. To protect our users, we can't process your request right now.
We'll restore your access as quickly as possible, so try again soon. In the meantime, if you suspect that your computer or network has been infected, you might want to run a virus checker or spyware remover to make sure that your systems are free of viruses and other spurious software.
If you're continually receiving this error, you may be able to resolve the problem by deleting your Google cookie and revisiting Google. For browser-specific instructions, please consult your browser's online support center.
We apologize for the inconvenience, and hope we'll see you again on Google.
To continue searching, please type the characters you see below: