

Why, What, How Practical introduction to SPARQL for biologists and informaticians

Using the real world UniProt and neXtProt
databases as illustrative examples



Swiss Institute of
Bioinformatics

Jerven Bolleman (Swiss-Prot)
Daniel Teixeira (CALIPHO)
Pierre-André Michel (CALIPHO)

neXtProt federated queries

Example with UniProt

find “Human proteins highly expressed in brain
and observed in a PDB struture involving a virus protein”

Other examples

<http://snorql.nextprot.org/>

Tag: *federated*

Example with UniProt

find “Human proteins highly expressed in brain
and observed in a PDB struture in complex with a virus protein”

SPARQL endpoint: <https://api.nextprot.org/sparql>

PREFIX ...

```
PREFIX up_core:<http://purl.uniprot.org/core/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
PREFIX up_database:<http://purl.uniprot.org/database/>
select distinct ?entry ?viralprotein ?db where {
{
  SERVICE <http://beta.sparql.uniprot.org/sparql> {
    SELECT distinct ?humanprotein ?viralprotein ?db WHERE {
      # get human proteins that share a PDB xref with a viral protein (same PDB id)
      ?humanprotein a up_core:Protein ; up_core:organism taxon:9606 .
      ?viralprotein a up_core:Protein ; up_core:organism / up_core:scientificName ?orgname.
      filter(contains(?orgname, "virus"))
      ?humanprotein rdfs:seeAlso ?db .
      ?viralprotein rdfs:seeAlso ?db .
      ?db up_core:database up_database:PDB .
    }
  }
}
# get human proteins highly expressed in brain according to HPA
?entry skos:exactMatch ?humanprotein .
?entry :isoform / :highExpression ?expressionprofile .
?expressionprofile :term / :childOf cv:TS-0095 .
?expressionprofile :evidence / :reference / :provenance db:HPA .
```

Example with UniProt

find “Human proteins highly expressed in brain
and observed in a PDB struture in complex with a virus protein”

SPARQL endpoint: <https://api.nextprot.org/sparql>

PREFIX ...

```
PREFIX up_core:<http://purl.uniprot.org/core/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
PREFIX up_database:<http://purl.uniprot.org/database/>
select distinct ?entry ?viralprotein ?db where {
  {
    SERVICE <http://beta.sparql.uniprot.org/sparql> {
      SELECT distinct ?humanprotein ?viralprotein ?db WHERE {
        # get human proteins that share a PDB xref with a viral protein (same PDB id)
        ?humanprotein a up_core:Protein ; up_core:organism taxon:9606 .
        ?viralprotein a up_core:Protein ; up_core:organism / up_core:scientificName ?orgname.
        filter(contains(?orgname, "virus"))
        ?humanprotein rdfs:seeAlso ?db .
        ?viralprotein rdfs:seeAlso ?db .
        ?db up_core:database up_database:PDB .
      }
    }
  }
  # get human proteins highly expressed in brain according to HPA
  ?entry skos:exactMatch ?humanprotein .
  ?entry :isoform / :highExpression ?expressionprofile .
  ?expressionprofile :term / :childOf cv:TS-0095 .
  ?expressionprofile :evidence / :reference / :provenance db:HPA .
}
```

Example with UniProt

find “Human proteins highly expressed in brain
and observed in a PDB struture in complex with a virus protein”

SPARQL endpoint: <https://api.nextprot.org/sparql>

PREFIX ...

```
PREFIX up_core:<http://purl.uniprot.org/core/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
PREFIX up_database:<http://purl.uniprot.org/database/>
select distinct ?entry ?viralprotein ?db where {
{
  SERVICE <http://beta.sparql.uniprot.org/sparql> {
    SELECT distinct ?humanprotein ?viralprotein ?db WHERE {
      # get human proteins that share a PDB xref with a viral protein (same PDB id)
      ?humanprotein a up_core:Protein ; up_core:organism taxon:9606 .
      ?viralprotein a up_core:Protein ; up_core:organism / up_core:scientificName ?orgname.
      filter(contains(?orgname, "virus"))
      ?humanprotein rdfs:seeAlso ?db .
      ?viralprotein rdfs:seeAlso ?db .
      ?db up_core:database up_database:PDB .
    }
  }
}
# get human proteins highly expressed in brain according to HPA
?entry skos:exactMatch ?humanprotein .
?entry :isoform / :highExpression ?expressionprofile .
?expressionprofile :term / :childOf cv:TS-0095 .
?expressionprofile :evidence / :reference / :provenance db:HPA .}
```

Example with UniProt

find “Human proteins highly expressed in brain
and observed in a PDB struture in complex with a virus protein”

SPARQL endpoint: <https://api.nextprot.org/sparql>

PREFIX ...

```
PREFIX up_core:<http://purl.uniprot.org/core/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
PREFIX up_database:<http://purl.uniprot.org/database/>
select distinct ?entry ?viralprotein ?db where {
  SERVICE <http://beta.sparql.uniprot.org/sparql> {
    SELECT distinct ?humanprotein ?viralprotein ?db WHERE {
      # get human proteins that share a PDB xref with a viral protein (same PDB id)
      ?humanprotein a up_core:Protein ; up_core:organism taxon:9606 .
      ?viralprotein a up_core:Protein ; up_core:organism / up_core:scientificName ?orgname .
      filter(contains(?orgname, "virus"))
      ?humanprotein rdfs:seeAlso ?db .
      ?viralprotein rdfs:seeAlso ?db .
      ?db up_core:database up_database:PDB .
    }
  }
}
# get human proteins highly expressed in brain according to HPA
?entry skos:exactMatch ?humanprotein .
?entry :isoform / :highExpression ?expressionprofile .
?expressionprofile :term / :childOf cv:TS-0095 .
?expressionprofile :evidence / :reference / :provenance db:HPA .
```

Example with UniProt

find “Human proteins highly expressed in brain
and observed in a PDB struture in complex with a virus protein”

SPARQL endpoint: <https://api.nextprot.org/sparql>

PREFIX ...

```
PREFIX up_core:<http://purl.uniprot.org/core/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
PREFIX up_database:<http://purl.uniprot.org/database/>
select distinct ?entry ?viralprotein ?db where {
{
  SERVICE <http://beta.sparql.uniprot.org/sparql> {
    SELECT distinct ?humanprotein ?viralprotein ?db WHERE {
      # get human proteins that share a PDB xref with a viral protein (same PDB id)
      ?humanprotein a up_core:Protein ; up_core:organism taxon:9606 .
      ?viralprotein a up_core:Protein ; up_core:organism / up_core:scientificName ?orgname .
      filter(contains(?orgname, "virus"))
      ?humanprotein rdfs:seeAlso ?db .
      ?viralprotein rdfs:seeAlso ?db .
      ?db up_core:database up_database:PDB .
    }
  }
}
# get human proteins highly expressed in brain according to HPA
?entry skos:exactMatch ?humanprotein .
?entry :isoform / :highExpression ?expressionprofile .
?expressionprofile :term / :childOf cv:TS-0095 .
?expressionprofile :evidence / :reference / :provenance db:HPA .
```

Example with UniProt

find “Human proteins highly expressed in brain
and observed in a PDB struture in complex with a virus protein”

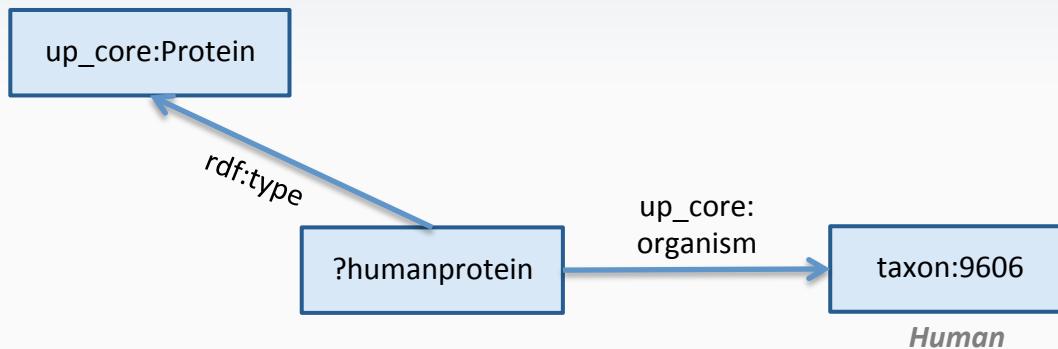
SPARQL endpoint: <https://api.nextprot.org/sparql>

PREFIX ...

```
PREFIX up_core:<http://purl.uniprot.org/core/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
PREFIX up_database:<http://purl.uniprot.org/database/>
select distinct ?entry ?viralprotein ?db where {
{
  SERVICE <http://beta.sparql.uniprot.org/sparql> {
    SELECT distinct ?humanprotein ?viralprotein ?db WHERE {
      # get human proteins that share a PDB xref with a viral protein (same PDB id)
      ?humanprotein a up_core:Protein ; up_core:organism taxon:9606 .
      ?viralprotein a up_core:Protein ; up_core:organism / up_core:scientificName ?orgname .
      filter(contains(?orgname, "virus"))
      ?humanprotein rdfs:seeAlso ?db .
      ?viralprotein rdfs:seeAlso ?db .
      ?db up_core:database up_database:PDB .
    }
  }
}
# get human proteins highly expressed in brain according to HPA
?entry skos:exactMatch ?humanprotein .
?entry :isoform / :highExpression ?expressionprofile .
?expressionprofile :term / :childOf cv:TS-0095 .
?expressionprofile :evidence / :reference / :provenance db:HPA .
```

Example with UniProt

find “Human proteins highly expressed in brain
and observed in a PDB struture in complex with a virus protein”



```
{\n    SERVICE <http://beta.sparql.uniprot.org/sparql> {\n        SELECT distinct ?humanprotein ?viralprotein ?db WHERE {\n            # get human proteins that share a PDB xref with a viral protein (same PDB id)\n            ?humanprotein a up_core:Protein ; up_core:organism taxon:9606 .\n            ?viralprotein a up_core:Protein ; up_core:organism / up_core:scientificName ?orgname .\n            filter(contains(?orgname, "virus"))\n            ?humanprotein rdfs:seeAlso ?db .\n            ?viralprotein rdfs:seeAlso ?db .\n            ?db up_core:database up_database:PDB .\n        }\n    }\n}
```

Example with UniProt

find “Human proteins highly expressed in brain
and observed in a PDB struture in complex with a virus protein”

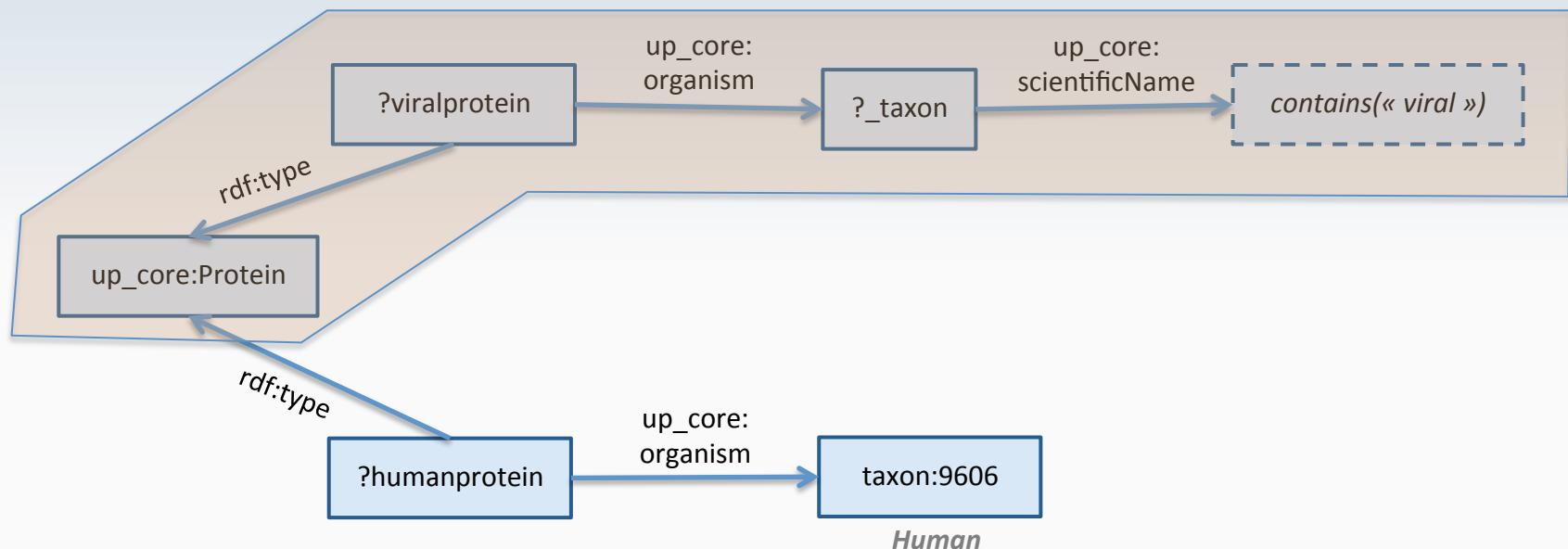
SPARQL endpoint: <https://api.nextprot.org/sparql>

PREFIX ...

```
PREFIX up_core:<http://purl.uniprot.org/core/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
PREFIX up_database:<http://purl.uniprot.org/database/>
select distinct ?entry ?viralprotein ?db where {
{
  SERVICE <http://beta.sparql.uniprot.org/sparql> {
    SELECT distinct ?humanprotein ?viralprotein ?db WHERE {
      # get human proteins that share a PDB xref with a viral protein (same PDB id)
      ?humanprotein a up_core:Protein ; up_core:organism taxon:9606 .
      ?viralprotein a up_core:Protein ; up_core:organism / up_core:scientificName ?orgname .
      filter(contains(?orgname, "virus"))
      ?humanprotein rdfs:seeAlso ?db .
      ?viralprotein rdfs:seeAlso ?db .
      ?db up_core:database up_database:PDB .
    }
  }
}
# get human proteins highly expressed in brain according to HPA
?entry skos:exactMatch ?humanprotein .
?entry :isoform / :highExpression ?expressionprofile .
?expressionprofile :term / :childOf cv:TS-0095 .
?expressionprofile :evidence / :reference / :provenance db:HPA .
}
```

Example with UniProt

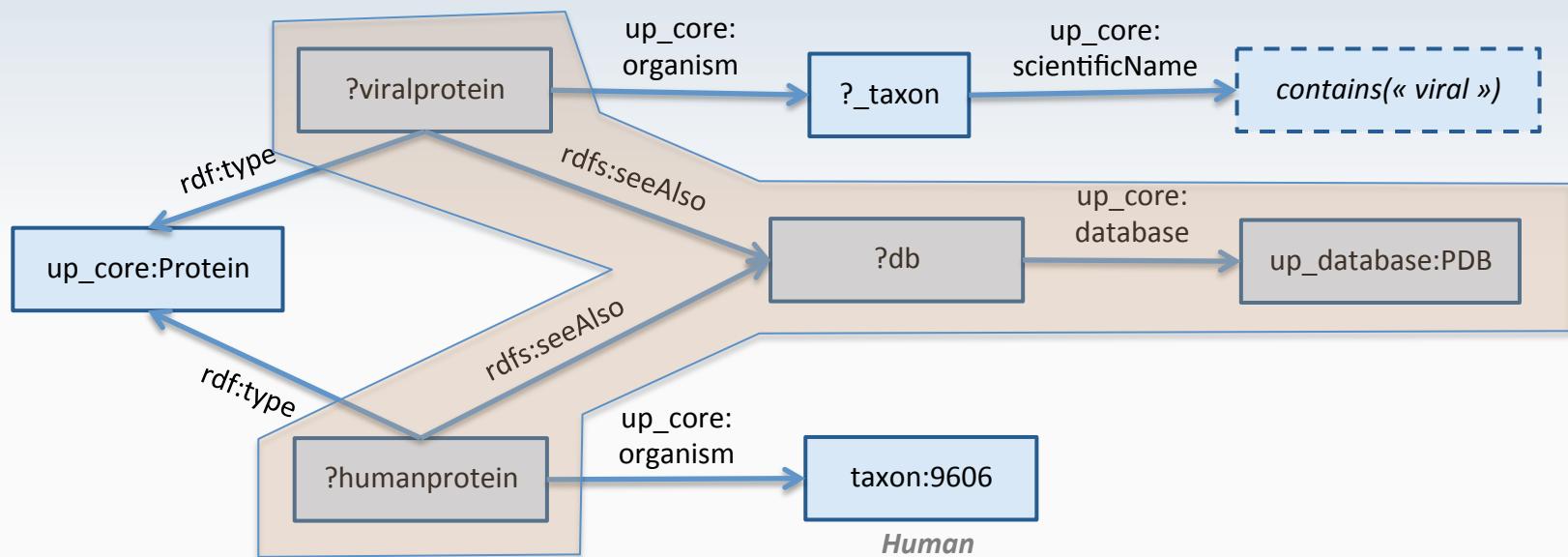
find “Human proteins highly expressed in brain
and observed in a PDB struture in complex with a virus protein”



```
{  
  SERVICE <http://beta.sparql.uniprot.org/sparql> {  
    SELECT distinct ?humanprotein ?viralprotein ?db WHERE {  
      # get human proteins that share a PDB xref with a viral protein (same PDB id)  
      ?humanprotein a up_core:Protein ; up_core:organism taxon:9606 .  
      ?viralprotein a up_core:Protein ; up_core:organism / up_core:scientificName ?orgname .  
      filter(contains(?orgname, "virus"))  
      ?humanprotein rdfs:seeAlso ?db .  
      ?viralprotein rdfs:seeAlso ?db .  
      ?db up_core:database up_database:PDB .  
    }  
  }  
}
```

Example with UniProt

find “Human proteins highly expressed in brain
and observed in a PDB struture in complex with a virus protein”



```

    {
        SERVICE <http://beta.sparql.uniprot.org/sparql> {
            SELECT distinct ?humanprotein ?viralprotein ?db WHERE {
                # get human proteins that share a PDB xref with a viral protein (same PDB id)
                ?humanprotein a up_core:Protein ; up_core:organism taxon:9606 .
                ?viralprotein a up_core:Protein ; up_core:organism / up_core:scientificName ?orgname .
                filter(contains(?orgname, "virus"))
                ?humanprotein rdfs:seeAlso ?db .
                ?viralprotein rdfs:seeAlso ?db .
                ?db up_core:database up_database:PDB .
            }
        }
    }
}
    
```

Example with UniProt

find “Human proteins highly expressed in brain
and observed in a PDB struture in complex with a virus protein”

SPARQL endpoint: <https://api.nextprot.org/sparql>

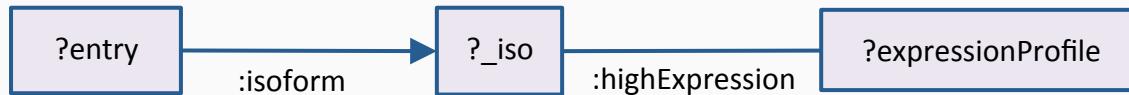
PREFIX ...

```
PREFIX up_core:<http://purl.uniprot.org/core/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
PREFIX up_database:<http://purl.uniprot.org/database/>
select distinct ?entry ?viralprotein ?db where {
{
  SERVICE <http://beta.sparql.uniprot.org/sparql> {
    SELECT distinct ?humanprotein ?viralprotein ?db WHERE {
      # get human proteins that share a PDB xref with a viral protein (same PDB id)
      ?humanprotein a up_core:Protein ; up_core:organism taxon:9606 .
      ?viralprotein a up_core:Protein ; up_core:organism / up_core:scientificName ?orgname.
      filter(contains(?orgname, "virus"))
      ?humanprotein rdfs:seeAlso ?db .
      ?viralprotein rdfs:seeAlso ?db .
      ?db up_core:database up_database:PDB .
    }
  }
}
# get human proteins highly expressed in brain according to HPA
?entry skos:exactMatch ?humanprotein .
?entry :isoform / :highExpression ?expressionprofile .
?expressionprofile :term / :childOf cv:TS-0095 .
?expressionprofile :evidence / :reference / :provenance db:HPA .
}
```

Example with UniProt

find “Human proteins highly expressed in brain
and observed in a PDB struture in complex with a virus protein”

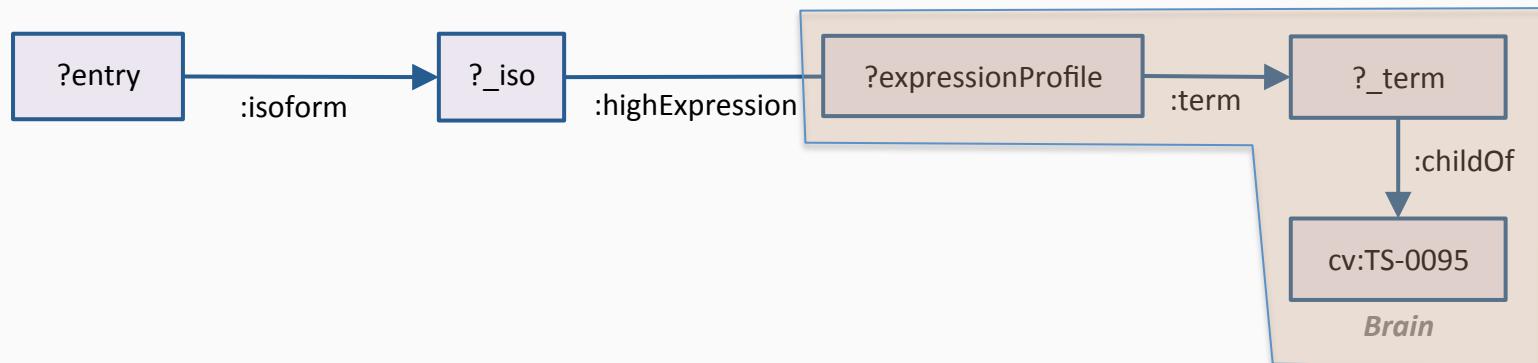
```
{  
  SERVICE <http://beta.sparql.uniprot.org/sparql> {  
    SELECT distinct ?humanprotein ?viralprotein ?db WHERE {  
      # get human proteins that share a PDB xref with a viral protein  
      ?humanprotein a up_core:Protein ; up_core:organism taxon:9606  
      ?viralprotein a up_core:Protein ; up_core:organism / up_core:s  
      filter(contains(?orgname, "virus"))  
      ?humanprotein rdfs:seeAlso ?db .  
      ?viralprotein rdfs:seeAlso ?db .  
      ?db up_core:database up_database:PDB .  
    }  
  }  
  # get human proteins highly expressed in brain according to HPA  
  ?entry skos:exactMatch ?humanprotein .  
  ?entry :isoform / :highExpression ?expressionprofile .  
  ?expressionprofile :term / :childOf cv:TS-0095 .  
  ?expressionprofile :evidence / :reference / :provenance db:HPA .  
}
```



Example with UniProt

find “Human proteins highly expressed in brain
and observed in a PDB struture in complex with a virus protein”

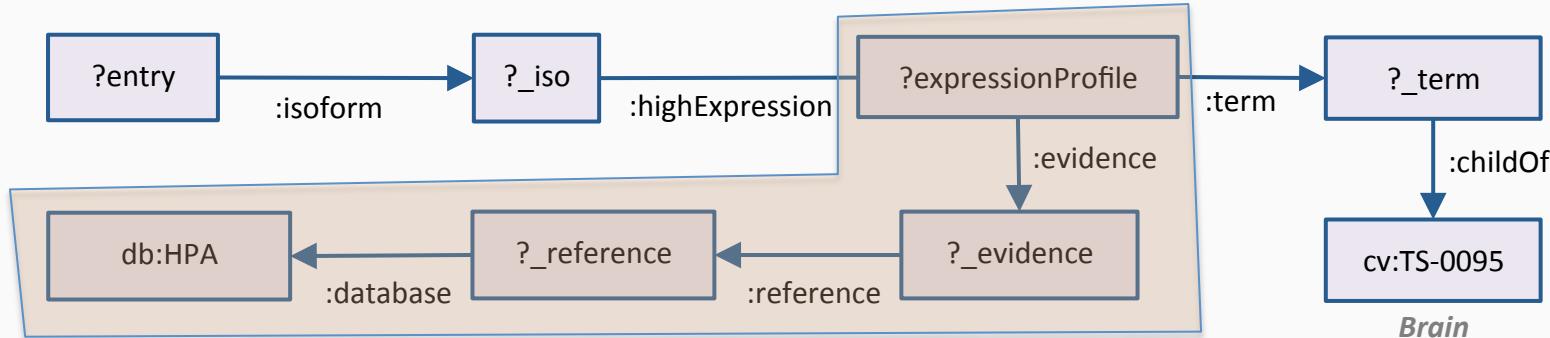
```
{  
  SERVICE <http://beta.sparql.uniprot.org/sparql> {  
    SELECT distinct ?humanprotein ?viralprotein ?db WHERE {  
      # get human proteins that share a PDB xref with a viral protein  
      ?humanprotein a up_core:Protein ; up_core:organism taxon:9606  
      ?viralprotein a up_core:Protein ; up_core:organism / up_core:s  
      filter(contains(?orgname, "virus"))  
      ?humanprotein rdfs:seeAlso ?db .  
      ?viralprotein rdfs:seeAlso ?db .  
      ?db up_core:database up_database:PDB .  
    }  
  }  
  # get human proteins highly expressed in brain according to HPA  
  ?entry skos:exactMatch ?humanprotein .  
  ?entry :isoform / :highExpression ?expressionprofile .  
  ?expressionprofile :term / :childOf cv:TS-0095 .  
  ?expressionprofile :evidence / :reference / :provenance db:HPA .  
}
```



Example with UniProt

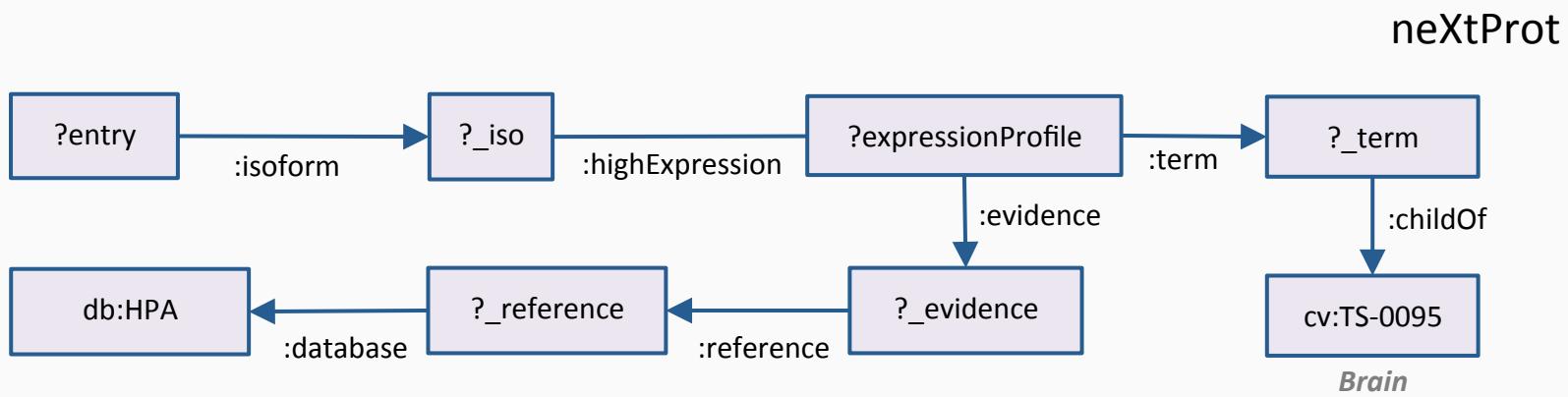
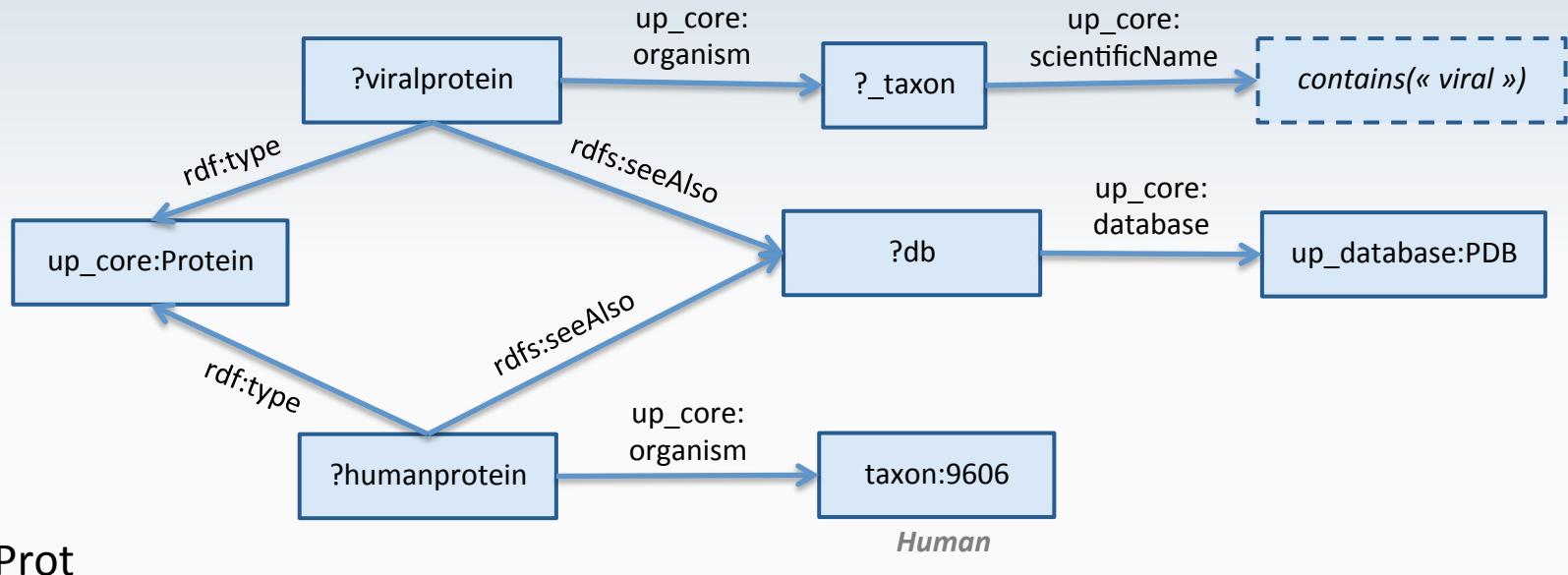
find “Human proteins highly expressed in brain
and observed in a PDB struture in complex with a virus protein”

```
{  
  SERVICE <http://beta.sparql.uniprot.org/sparql> {  
    SELECT distinct ?humanprotein ?viralprotein ?db WHERE {  
      # get human proteins that share a PDB xref with a viral protein  
      ?humanprotein a up_core:Protein ; up_core:organism taxon:9606  
      ?viralprotein a up_core:Protein ; up_core:organism / up_core:s  
      filter(contains(?orgname, "virus"))  
      ?humanprotein rdfs:seeAlso ?db .  
      ?viralprotein rdfs:seeAlso ?db .  
      ?db up_core:database up_database:PDB .  
    }  
  }  
  # get human proteins highly expressed in brain according to HPA  
  ?entry skos:exactMatch ?humanprotein .  
  ?entry :isoform / :highExpression ?expressionprofile .  
  ?expressionprofile :term / :childOf cv:TS-0095 .  
  ?expressionprofile :evidence / :reference / :provenance db:HPA .  
}
```



Example with UniProt

find “Human proteins highly expressed in brain
and observed in a PDB struture in complex with a virus protein”



Example with UniProt

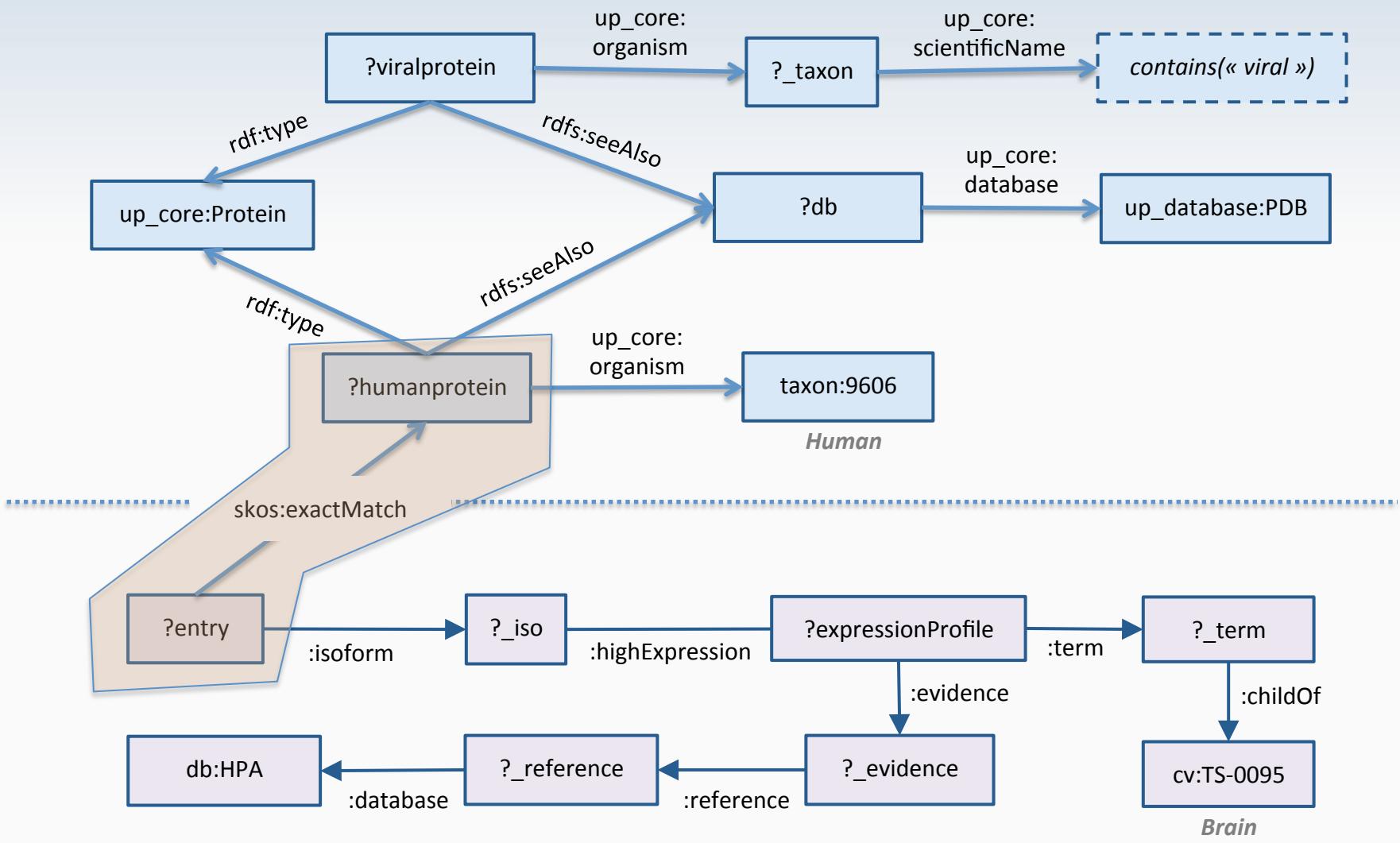
find “Human proteins highly expressed in brain
and observed in a PDB struture in complex with a virus protein”

SPARQL endpoint: <https://api.nextprot.org/sparql>

PREFIX ...

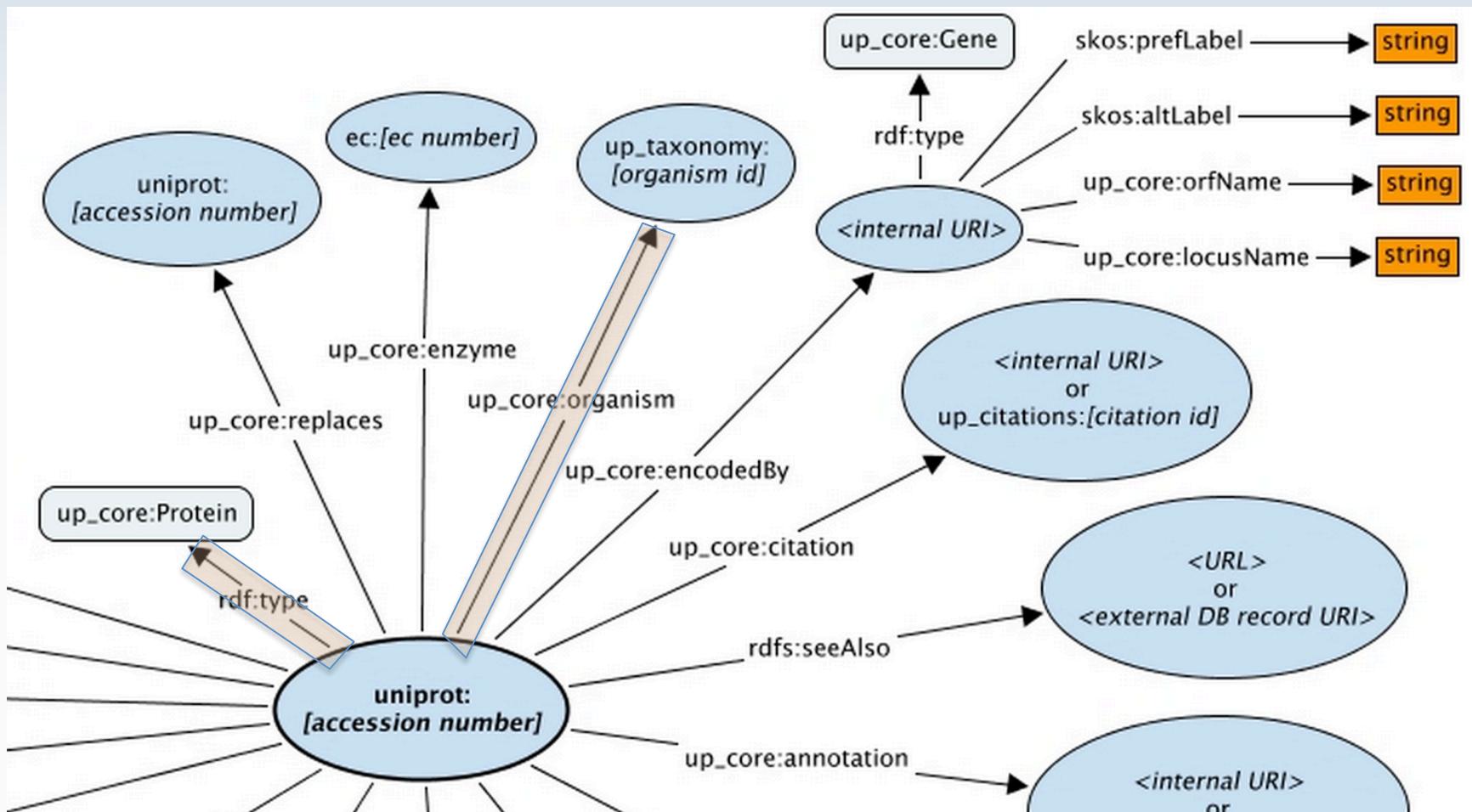
```
PREFIX up_core:<http://purl.uniprot.org/core/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
PREFIX up_database:<http://purl.uniprot.org/database/>
select distinct ?entry ?viralprotein ?db where {
{
  SERVICE <http://beta.sparql.uniprot.org/sparql> {
    SELECT distinct ?humanprotein ?viralprotein ?db WHERE {
      # get human proteins that share a PDB xref with a viral protein (same PDB id)
      ?humanprotein a up_core:Protein ; up_core:organism taxon:9606 .
      ?viralprotein a up_core:Protein ; up_core:organism / up_core:scientificName ?orgname.
      filter(contains(?orgname, "virus"))
      ?humanprotein rdfs:seeAlso ?db .
      ?viralprotein rdfs:seeAlso ?db .
      ?db up_core:database up_database:PDB .
    }
  }
}
# get human proteins highly expressed in brain according to HPA
?entry skos:exactMatch ?humanprotein .
?entry :isoform / :highExpression ?expressionprofile .
?expressionprofile :term / :childOf cv:TS-0095 .
?expressionprofile :evidence / :reference / :provenance db:HPA .
}
```

Full query graph



Example with UniProt

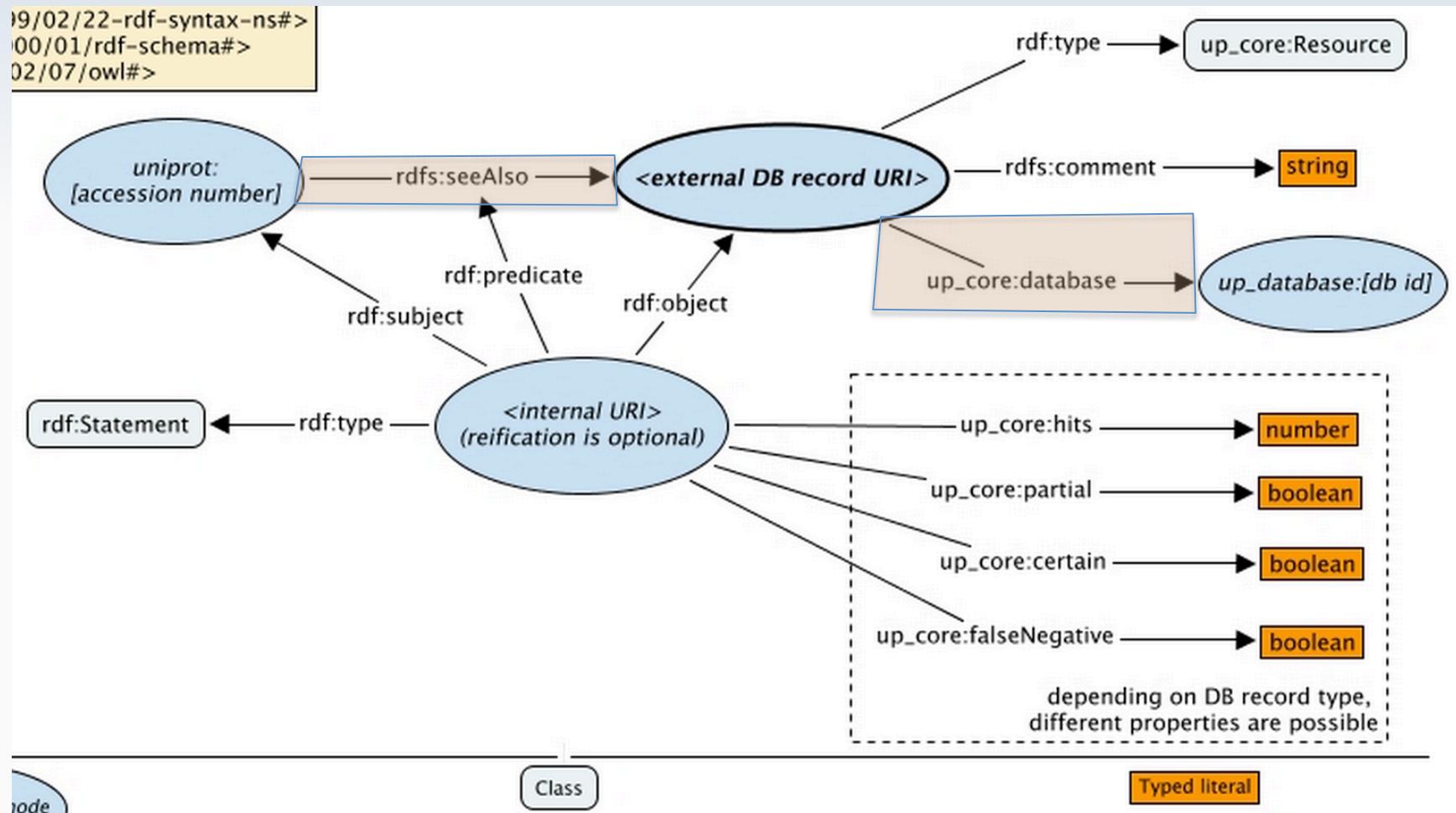
Documentation – UniProt graph



<http://beta.sparql.uniprot.org/uniprot>

Example with UniProt

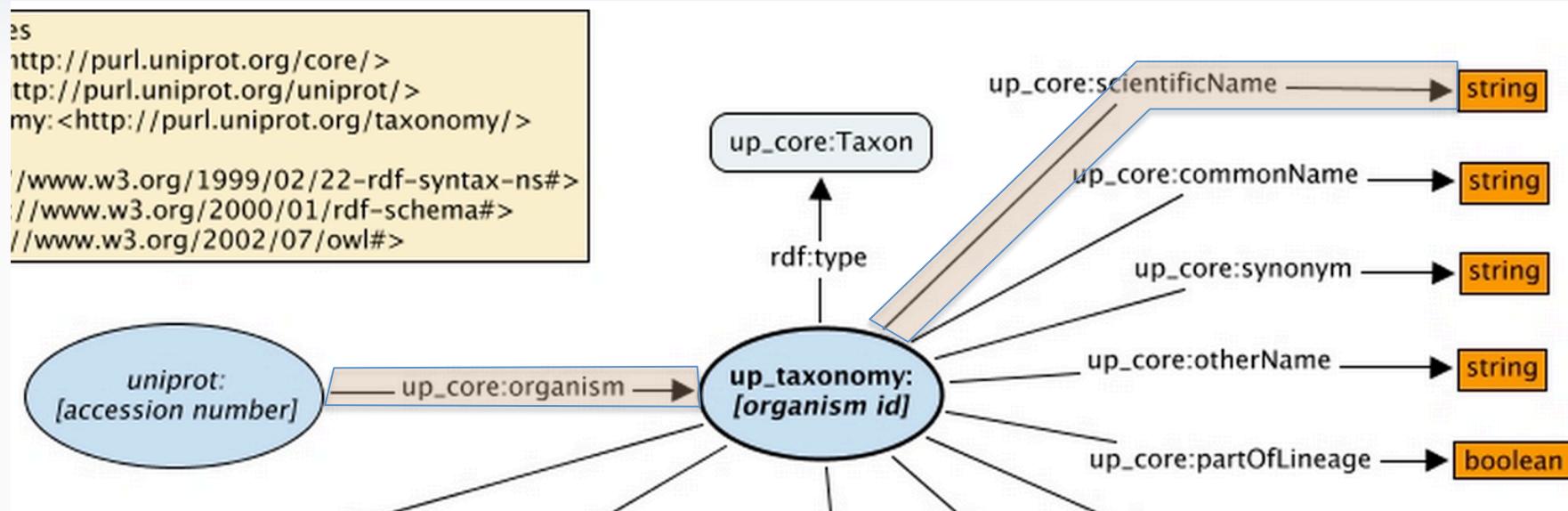
Documentation – UniProt graph & cross references



<http://beta.sparql.uniprot.org/uniprot>

Example with UniProt

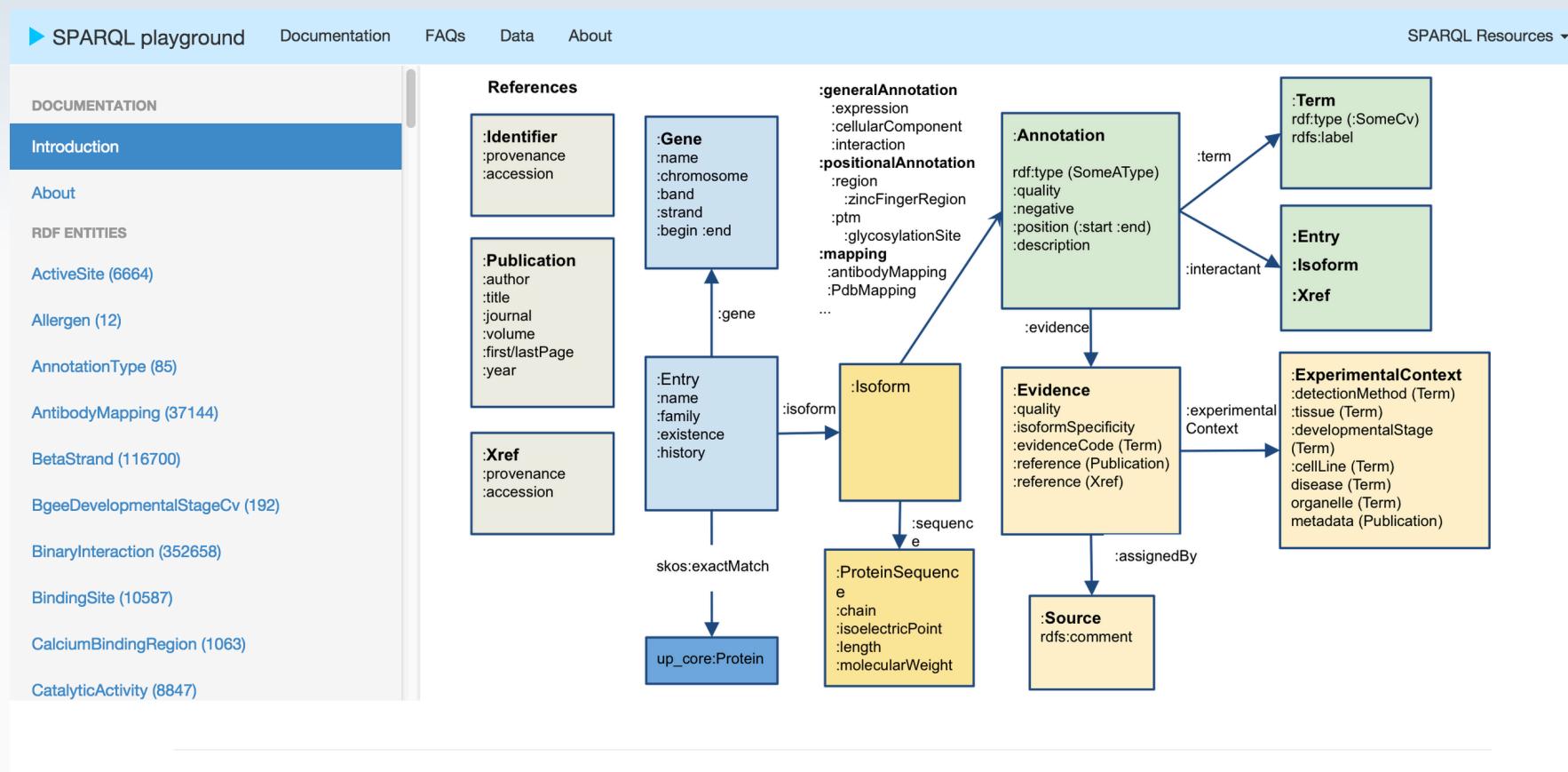
Documentation – Taxonomy graph



<http://beta.sparql.uniprot.org/taxonomy>

Example with UniProt

Documentation – neXtProt data model



Playground: <http://sparql-playground.nextprot.org/help/doc/introduction>
 Production: <https://search.nextprot.org/help/learn-advanced-search>

Example with UniProt

Documentation – neXtProt data model browser

The screenshot shows the neXtProt data model browser interface. The top navigation bar includes links for SPARQL playground, Documentation, FAQs, Data, and About. A sidebar on the left lists various entities with their counts: SignalPeptide (6009), SmallMoleculeInteraction (24952), Source (98), SrmPeptideMapping (295209), SubcellularLocation (85900) which is highlighted in blue, SubcellularLocationNote (11342), TopologicalDomain (30681), TransmembraneRegion (38054), Turn (27153), UnipathwayCv (2559), UniprotDiseaseCv (4145), UniprotFamilyCv (9792), and UniprotKeyword (520558). The main content area displays the 'SubcellularLocation' entity. It features a summary box with the title 'SubcellularLocation' (85900), a 'Values' dropdown, and a count of 85900. Below this is a description: 'Subcellular location of the mature protein'. A list of predicates associated with this entity includes ':Isoform :cellularComponent', ':Isoform :generalAnnotation', and ':Isoform :subcellularLocation'. To the right, there is a detailed view of the ':SubcellularLocation' resource, showing its properties: ':entryAnnotationId' (xsd:string, 85900), ':evidence' (Evidence, 107145), ':quality' (QualityQualifier, 85900), ':term' (UniprotSubcellularLocationCv, 85900), and 'rdfs:comment' (xsd:string, 85900). A note below states 'A band'. At the bottom, there is an 'Example' section with the following SPARQL query results:

```
annotation:NX_A0JNW5-1-AN_A0JNW5_000439 :entryAnnotationId "AN_A0JNW5_000439" .  
annotation:NX_A0PK00-1-AN_A0PK00_000854 :evidence evidence:evidence:84048336 .  
annotation:NX_A0PK00-1-AN_A0PK00_0001 :quality :GOLD .
```

Playground: <http://sparql-playground.nextprot.org/help/entity/SubcellularLocation>
Production: <http://snorql.nextprot.org/help/entity/SubcellularLocation>

Why use RDF / SPARQL

- Based on W3C standards
- Framework for both data & ontologies
- Self explanatory
- Shared URIs
- Query expressivity, precise answers

The end

Many thanks to

Jerven Bolleman

Daniel Teixeira

Alain Gateau

Monique Zahn

Pascale Gaudet