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

Using the real world UniProt and neXtProt databases as illustrative examples

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

SIB
Swiss Institute of Bioinformatics

elixir SWITZERLAND

neXtProt and SPARQL

SPARQL as the advanced search system

- Integrated in UI
- More than 100 sample queries
- Extension of SNORQL as a toolbox for users to work out queries
- Help service
- Persistence of user queries

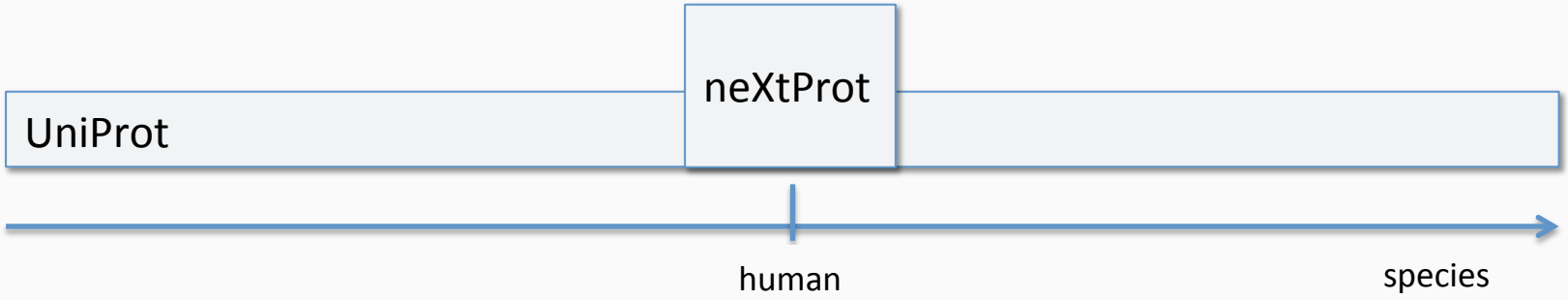
neXtProt content

Specificity

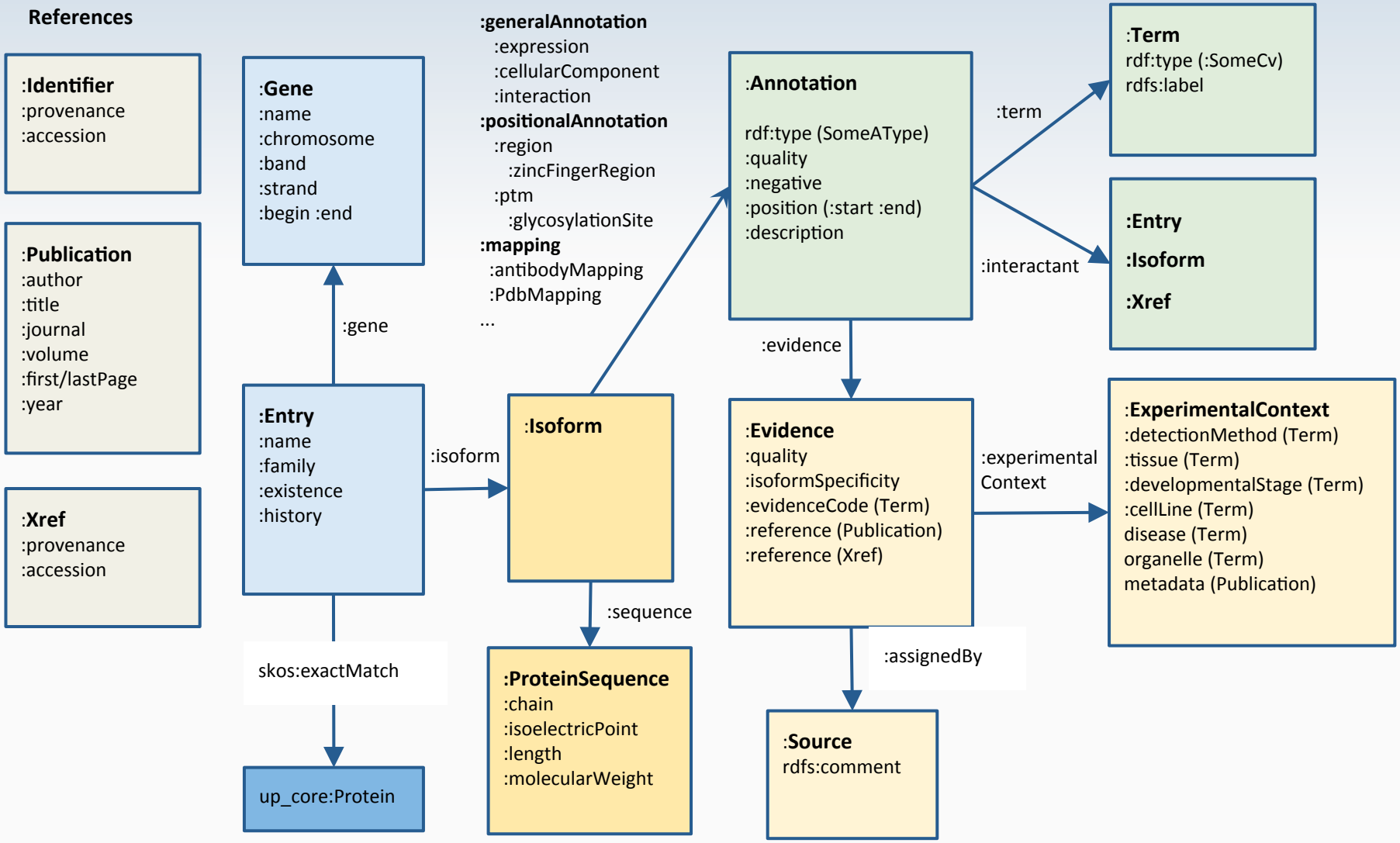
- Human centric
- Isoform centric

Data sources

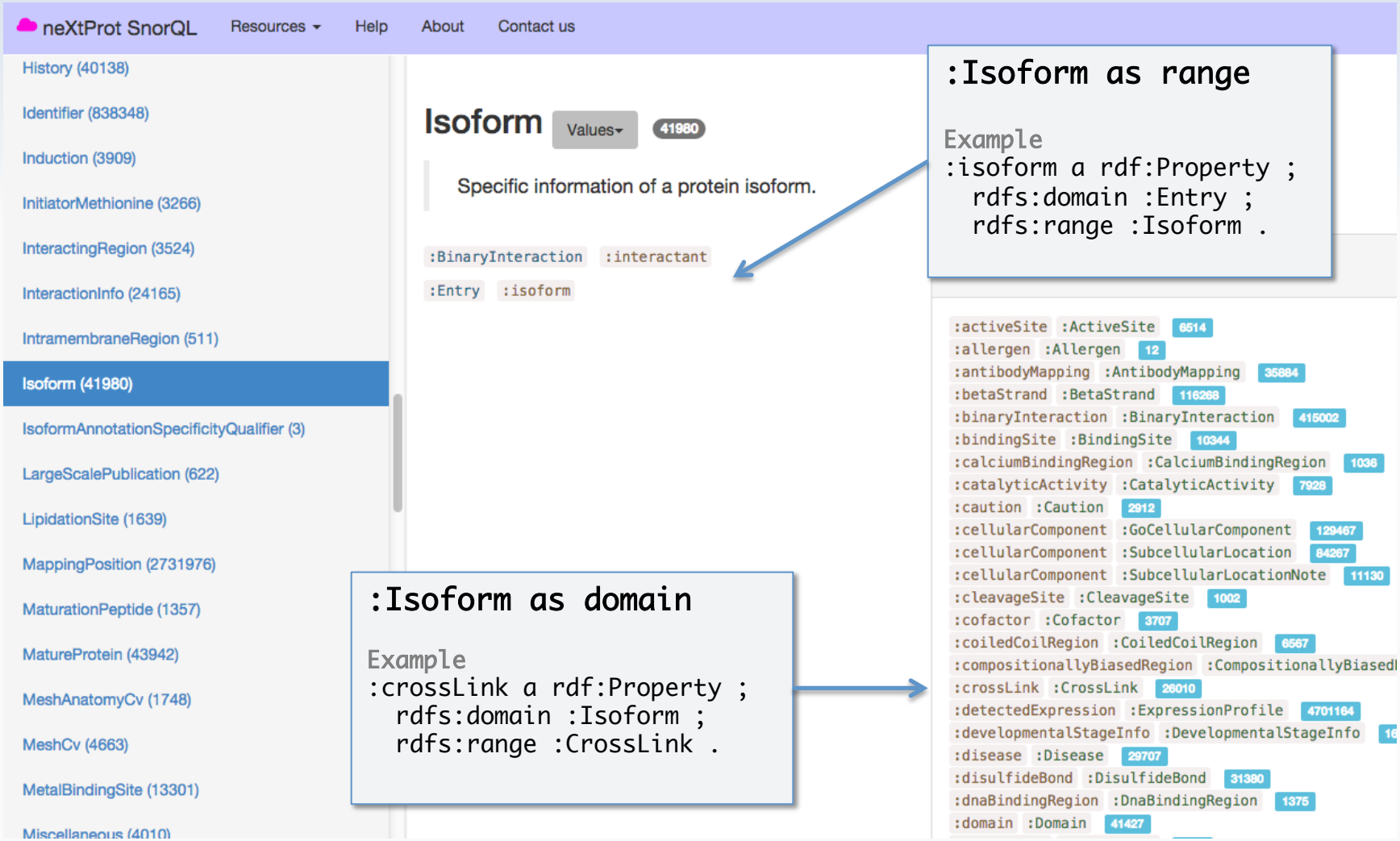
- UniProt
- proteomics: PeptideAtlas, SRM atlas
- localization: DKFZ, DYP, GO
- variants: UniProtKB, dbSNP via Ensembl, Cosmic, Clinvar
- expression: BGee, HPA
- function: full set of human GO annotations
- interactions: IntAct silver quality interactions



neXtProt - RDF model overview



neXtProt –data model browser



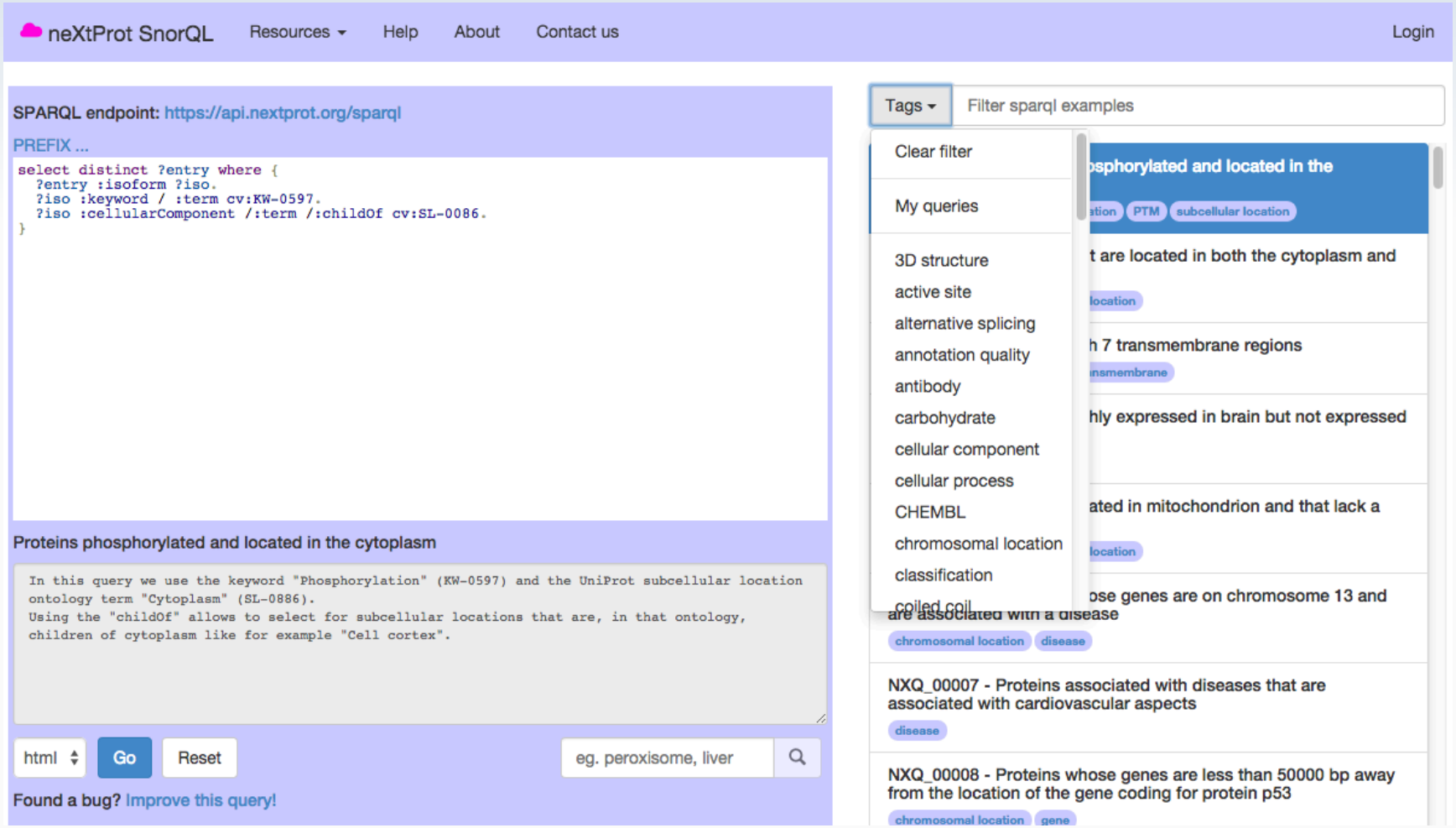
Example

Looking for

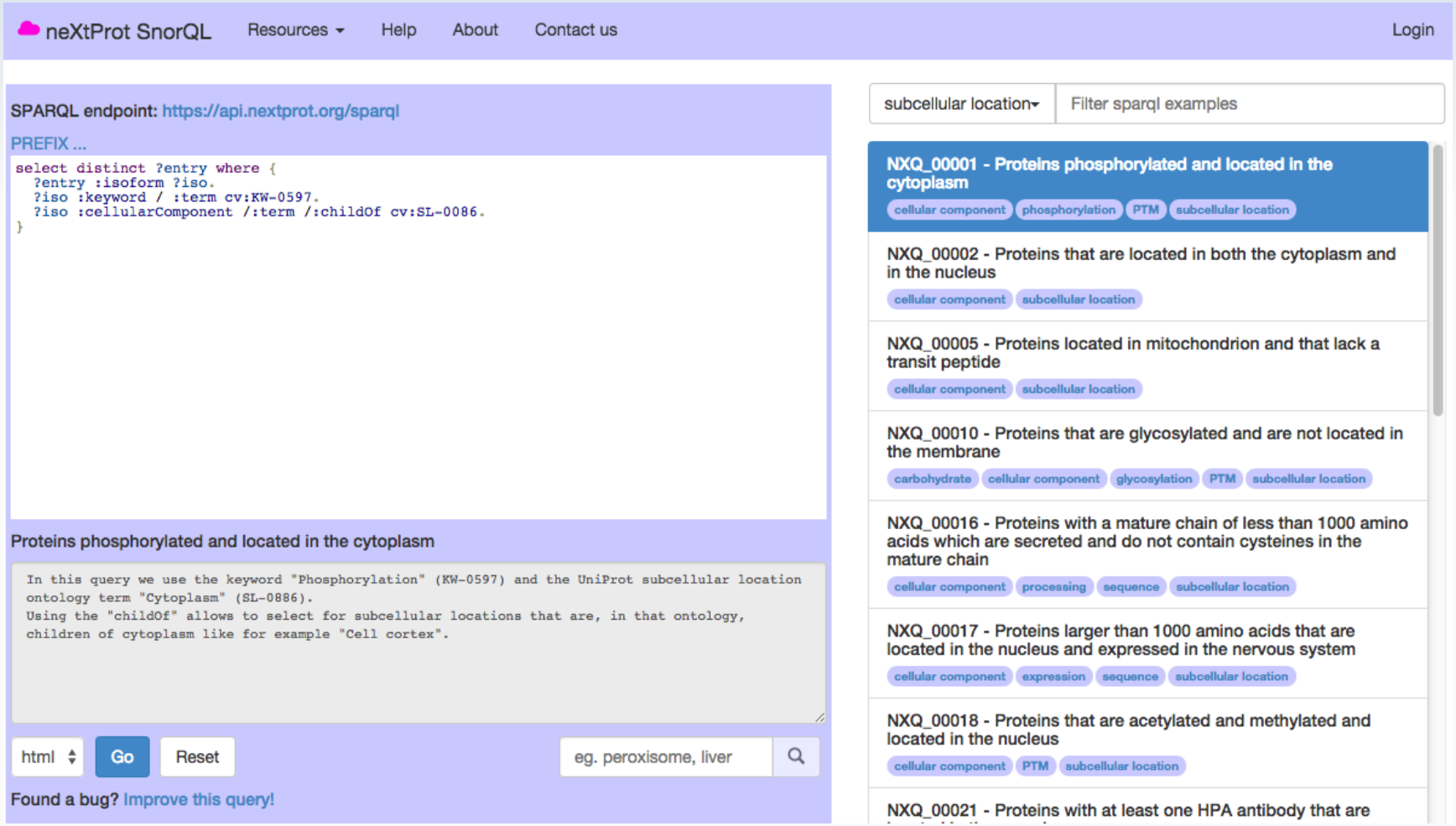
"phosphorylated proteins located in the Golgi apparatus"

- Search the existing sample queries
- Understand the query best matching your needs
- Modify it or get help
- Run your query
- Save it for reuse (not in SPARQL playground)

neXtProt – searching samples queries



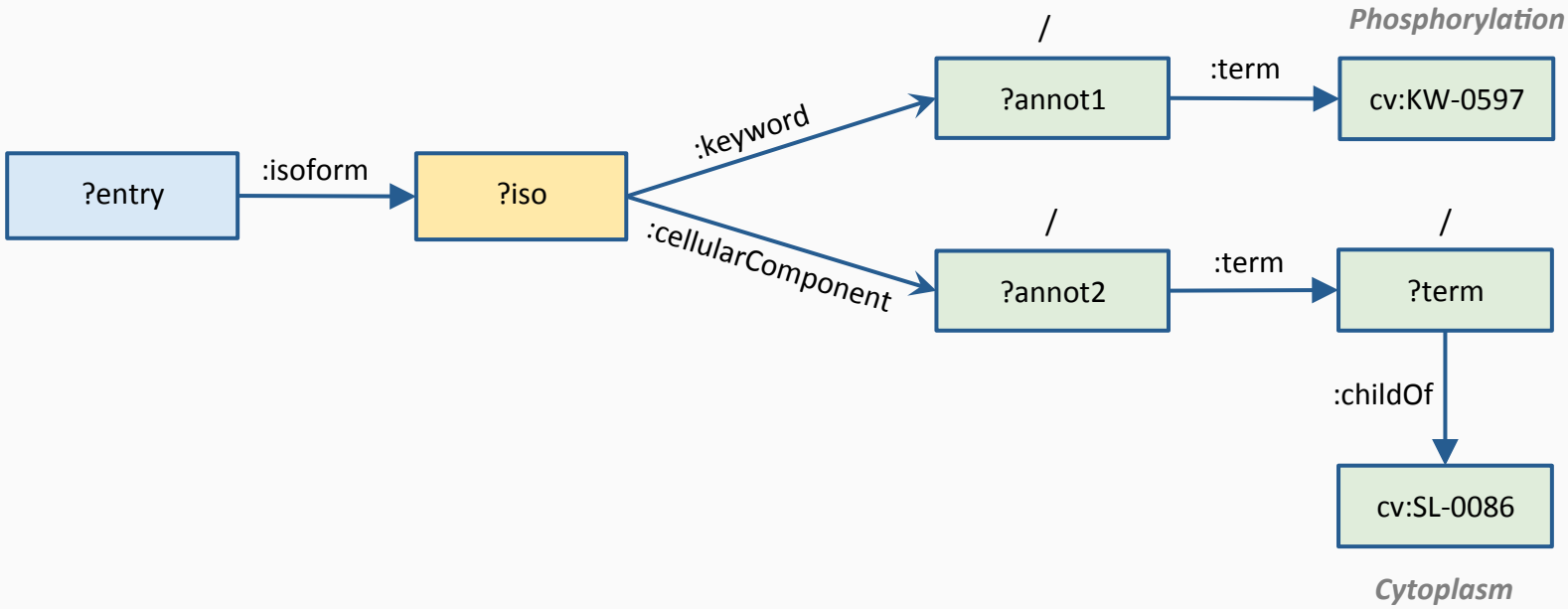
neXtProt – selecting a sample query



neXtProt – understanding the query

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

```
PREFIX ...
select distinct ?entry where {
  ?entry :isoform ?iso.
  ?iso :keyword / :term cv:KW-0597.
  ?iso :cellularComponent /:term /:childOf cv:SL-0086.
}
```

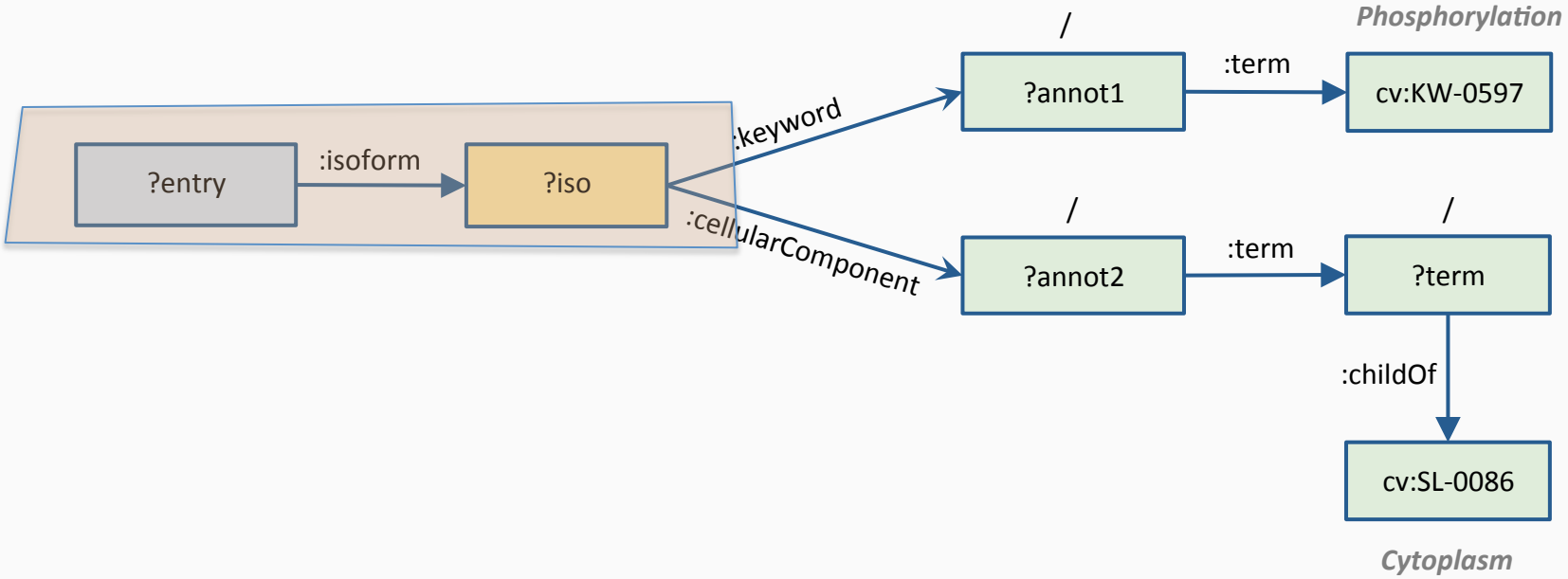


neXtProt – understanding the query

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

PREFIX ...

```
select distinct ?entry where {
  ?entry :isoform ?iso.
  ?iso :keyword / :term cv:KW-0597.
  ?iso :cellularComponent /:term /:childOf cv:SL-0086.
}
```

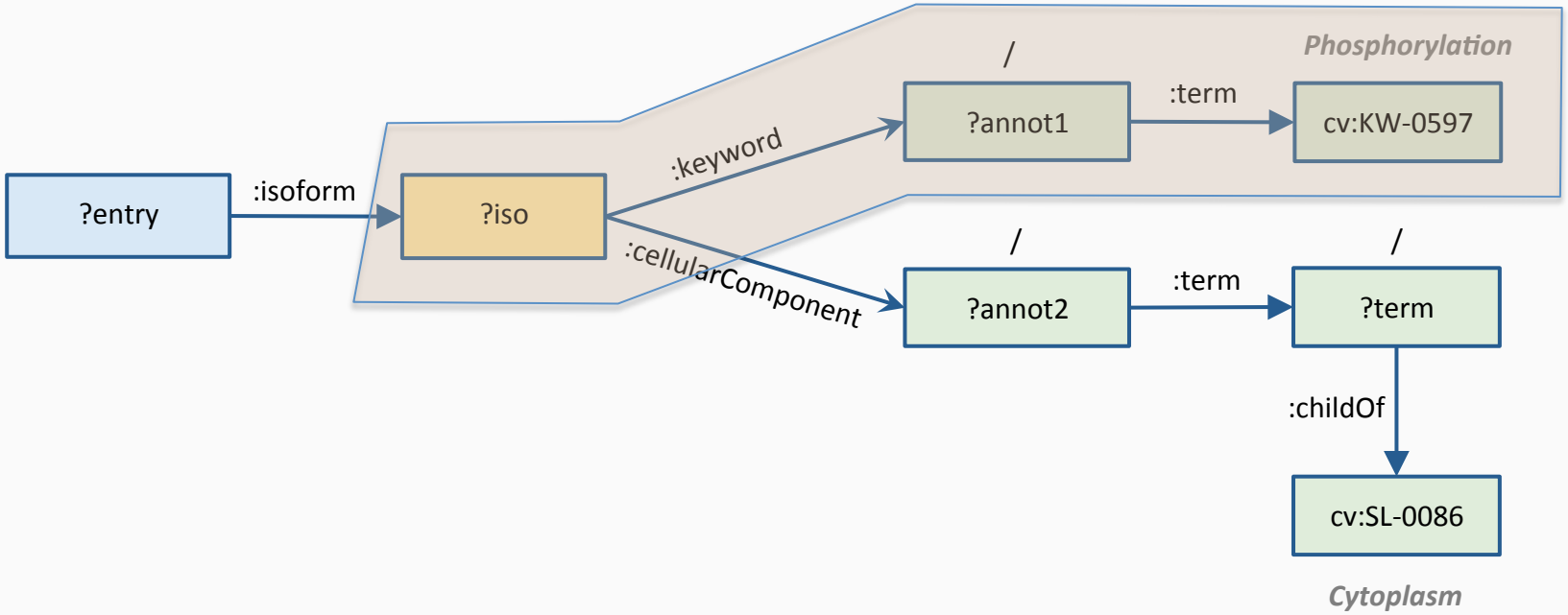


neXtProt – understanding the query

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

PREFIX ...

```
select distinct ?entry where {
  ?entry :isoform ?iso.
  ?iso :keyword / :term cv:KW-0597.
  ?iso :cellularComponent /:term /:childOf cv:SL-0086.
}
```

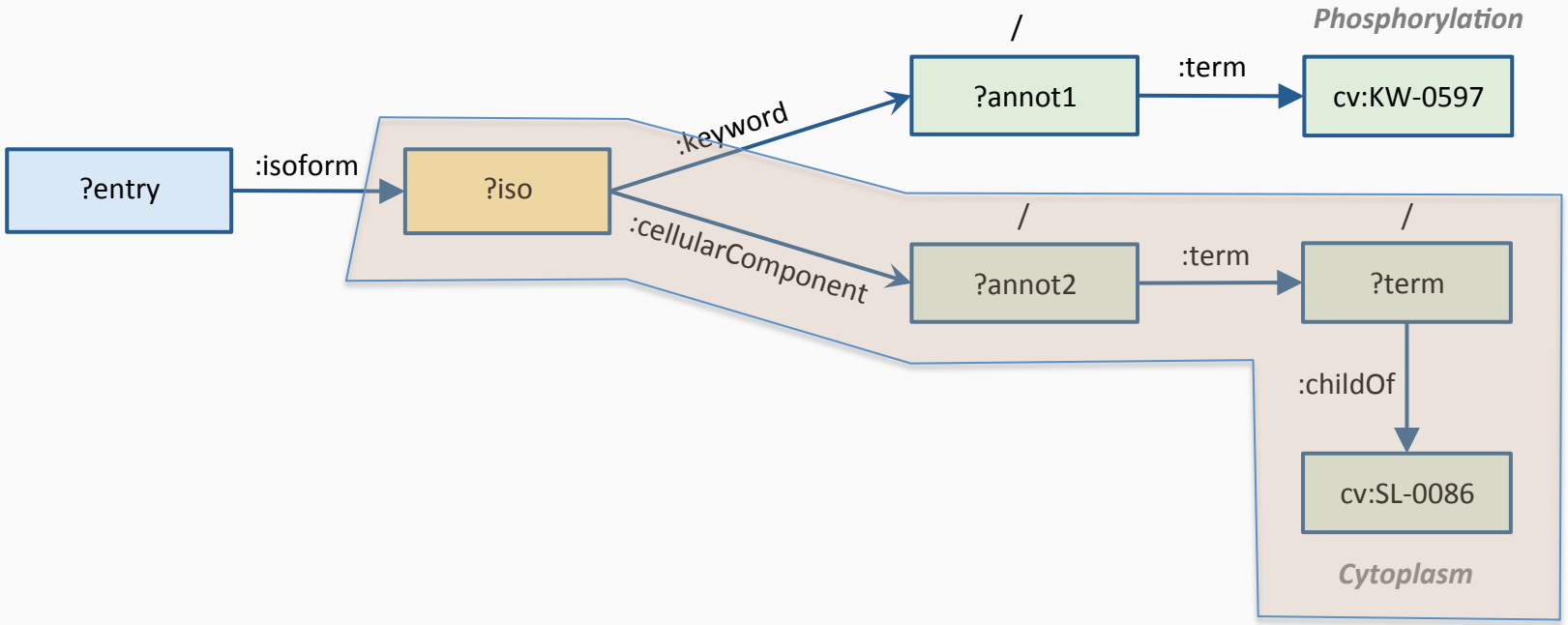


neXtProt – understanding the query

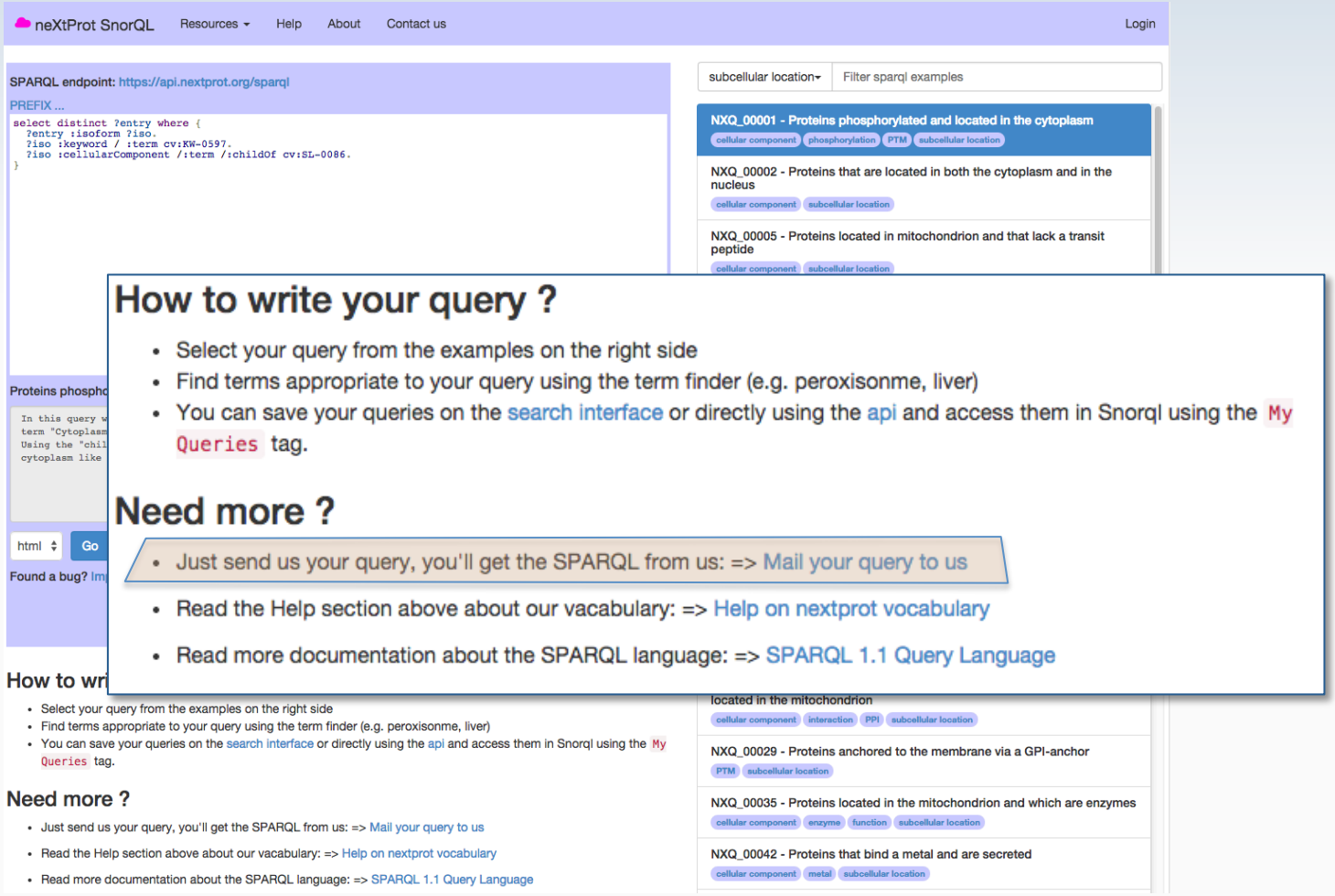
SPARQL endpoint: https://api.nextprot.org/sparql

PREFIX ...

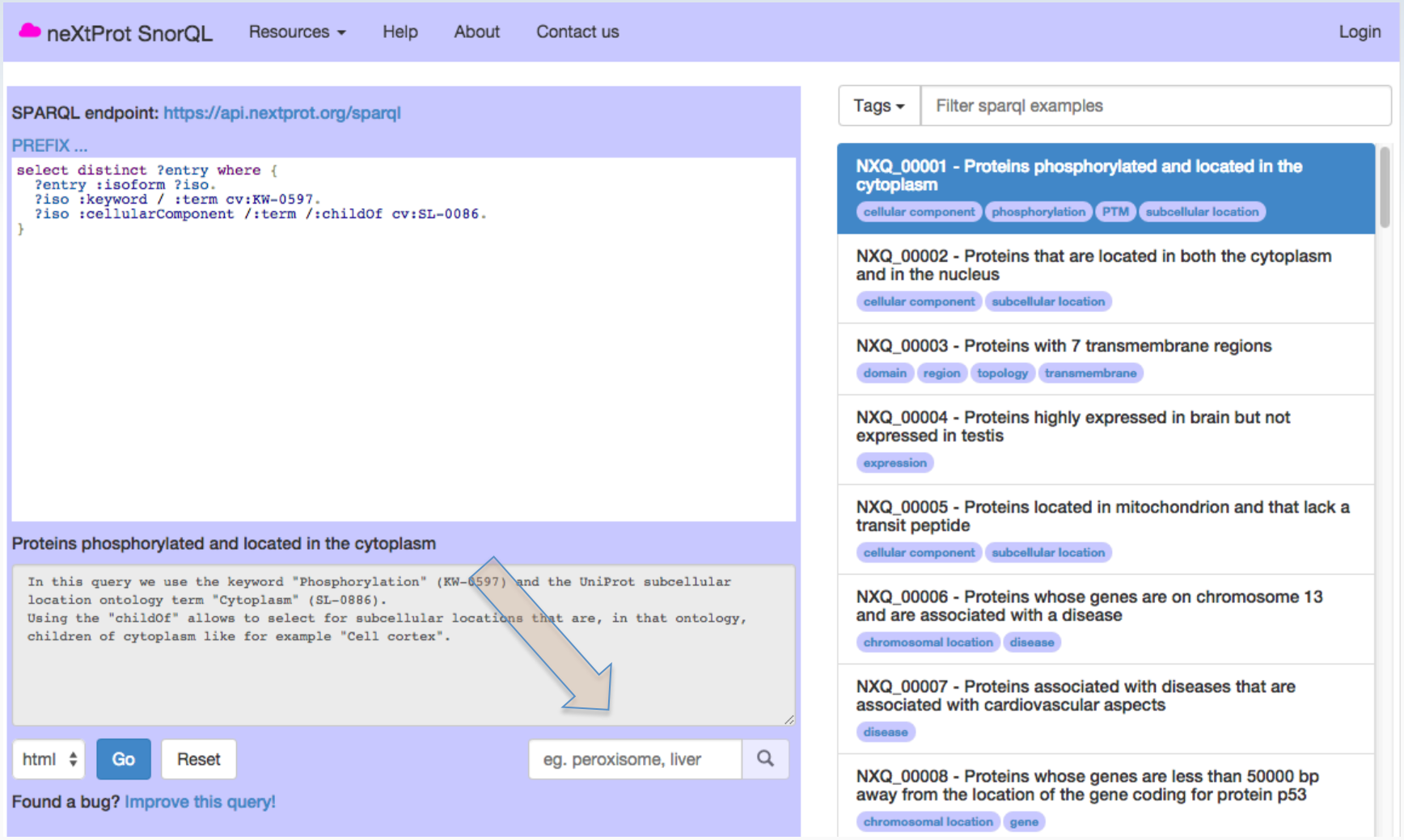
```
select distinct ?entry where {
  ?entry :isoform ?iso.
  ?iso :keyword / :term cv:KW-0597.
  ?iso :cellularComponent /:term /:childOf cv:SL-0086.
}
```



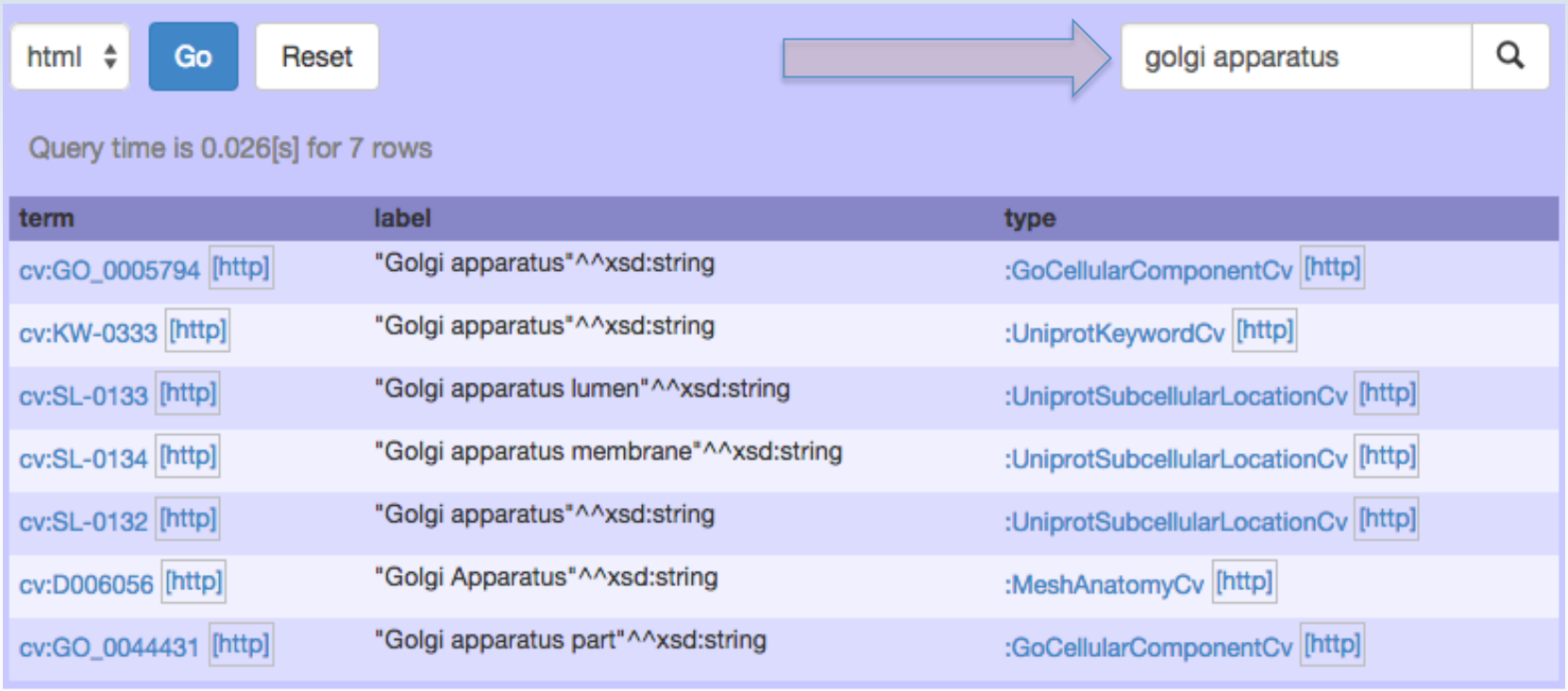
neXtProt – I am lost, help !



neXtProt – searching terms

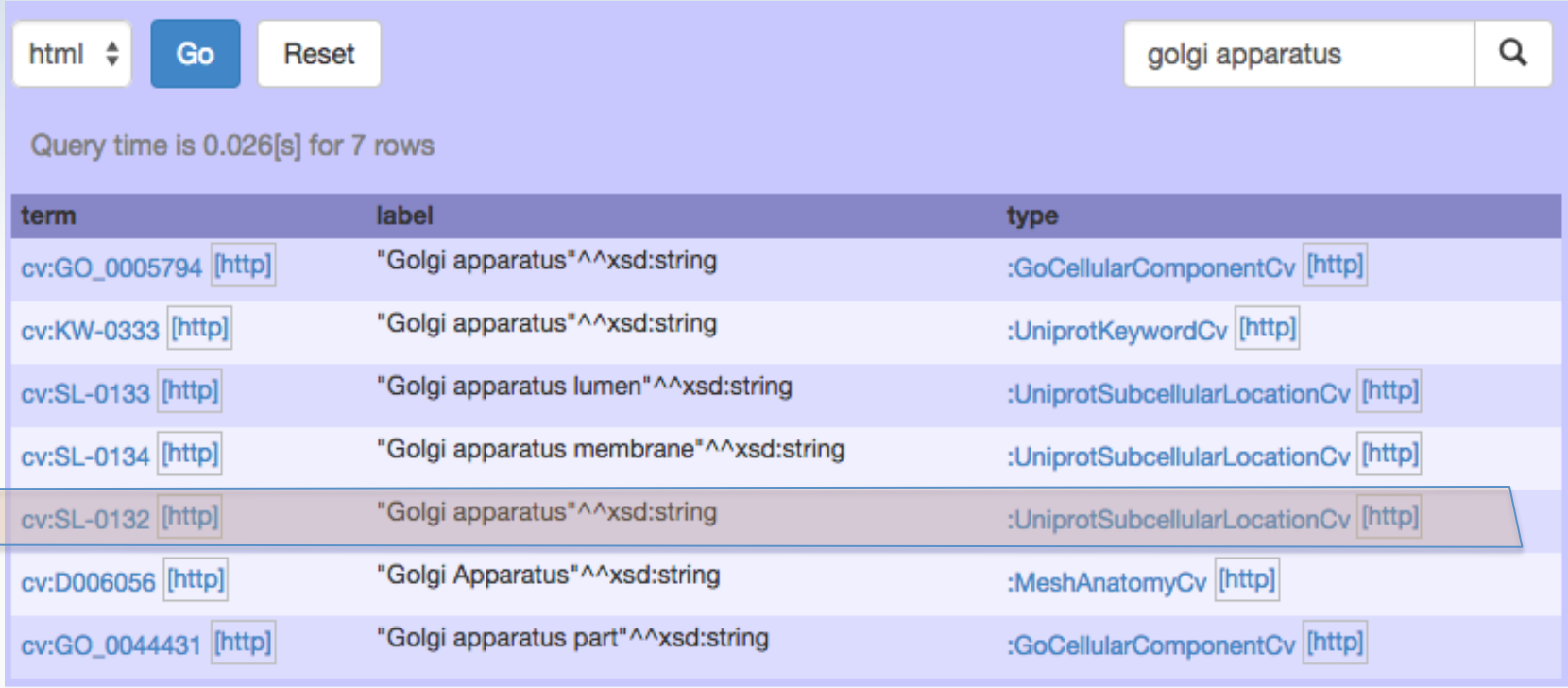


neXtProt – searching terms



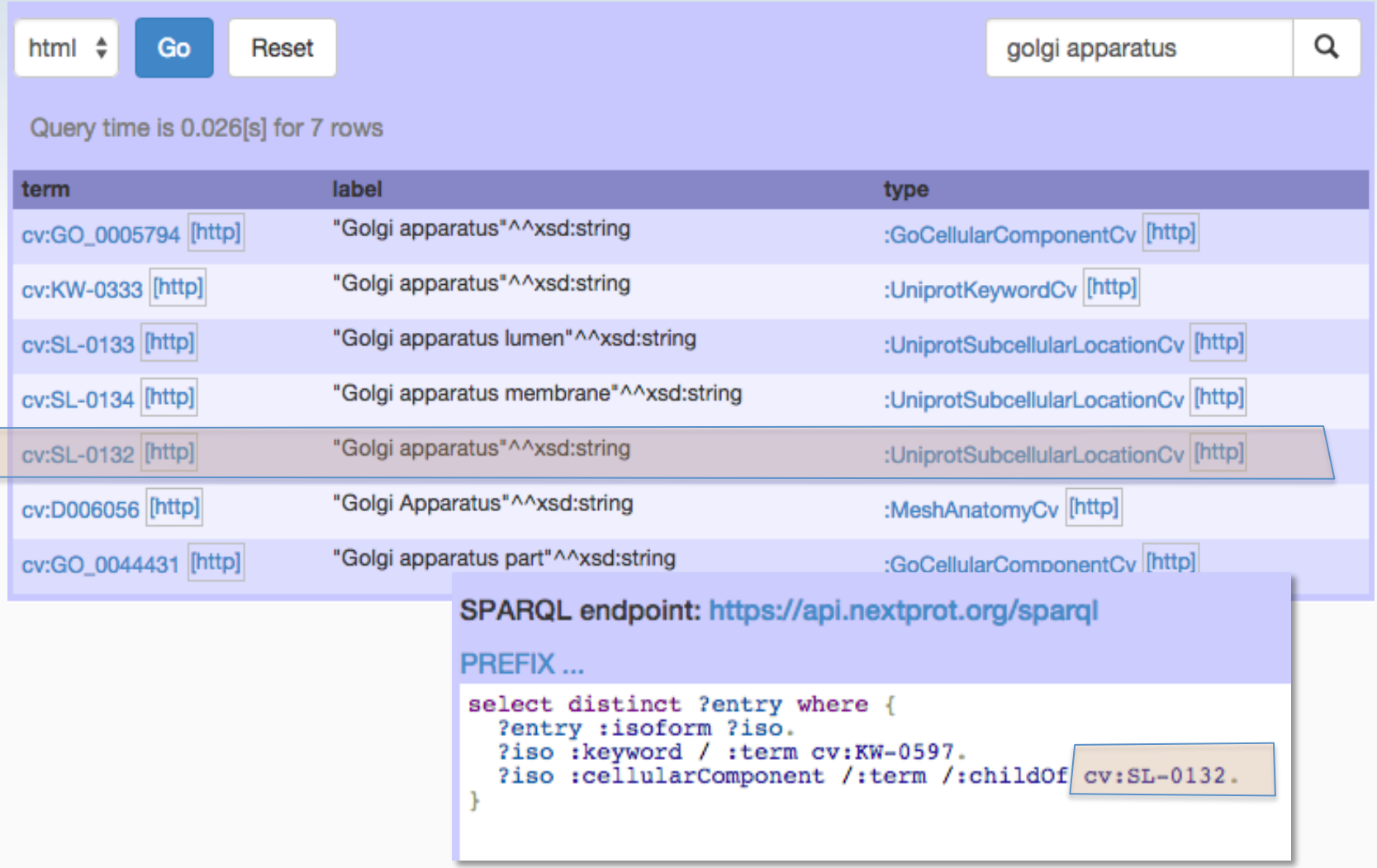
term	label	type
cv:GO_0005794 [http]	"Golgi apparatus"^^xsd:string	:GoCellularComponentCv [http]
cv:KW-0333 [http]	"Golgi apparatus"^^xsd:string	:UniprotKeywordCv [http]
cv:SL-0133 [http]	"Golgi apparatus lumen"^^xsd:string	:UniprotSubcellularLocationCv [http]
cv:SL-0134 [http]	"Golgi apparatus membrane"^^xsd:string	:UniprotSubcellularLocationCv [http]
cv:SL-0132 [http]	"Golgi apparatus"^^xsd:string	:UniprotSubcellularLocationCv [http]
cv:D006056 [http]	"Golgi Apparatus"^^xsd:string	:MeshAnatomyCv [http]
cv:GO_0044431 [http]	"Golgi apparatus part"^^xsd:string	:GoCellularComponentCv [http]

neXtProt – searching terms



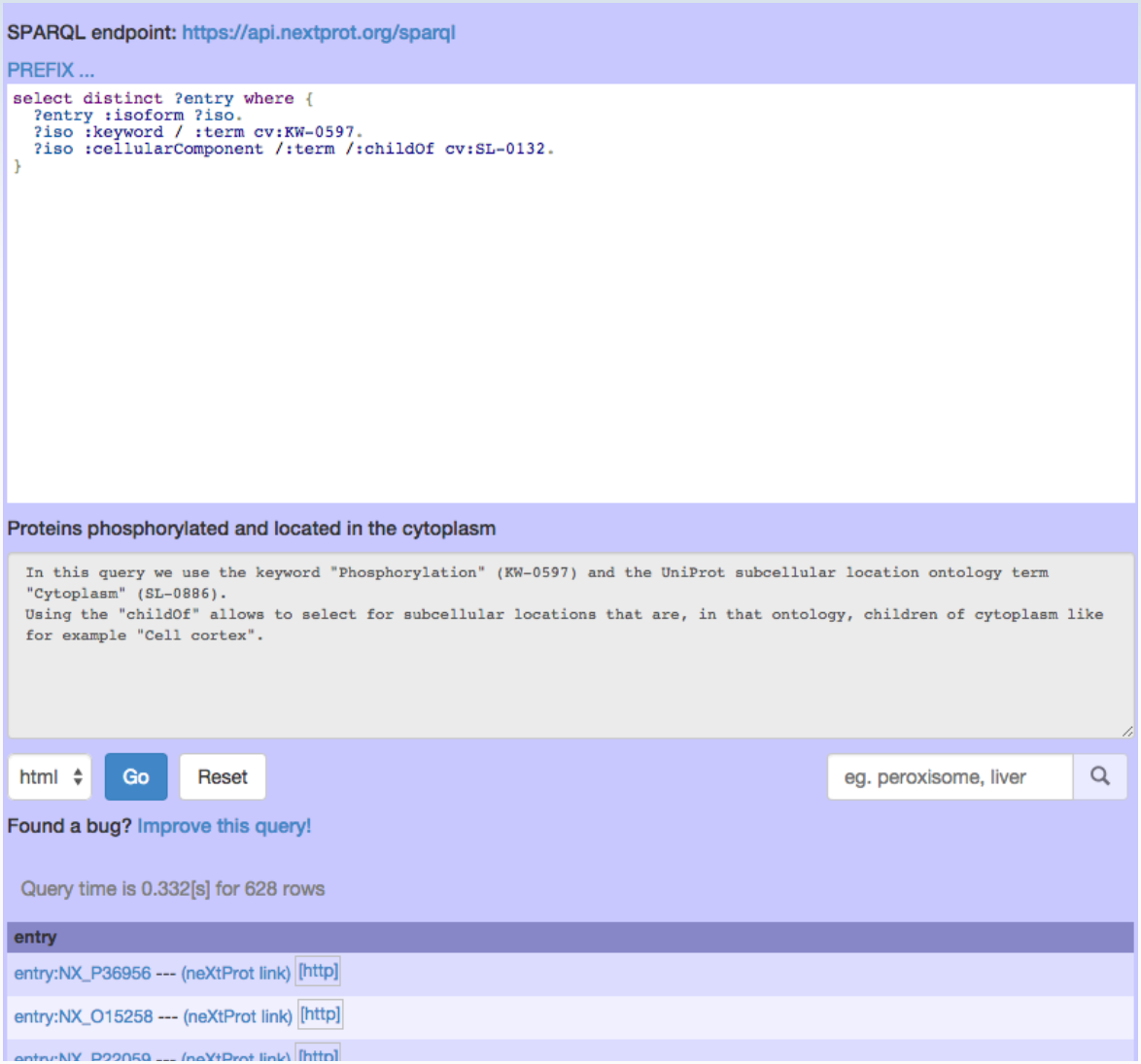
term	label	type
cv:GO_0005794 [http]	"Golgi apparatus"^^xsd:string	:GoCellularComponentCv [http]
cv:KW-0333 [http]	"Golgi apparatus"^^xsd:string	:UniprotKeywordCv [http]
cv:SL-0133 [http]	"Golgi apparatus lumen"^^xsd:string	:UniprotSubcellularLocationCv [http]
cv:SL-0134 [http]	"Golgi apparatus membrane"^^xsd:string	:UniprotSubcellularLocationCv [http]
cv:SL-0132 [http]	"Golgi apparatus"^^xsd:string	:UniprotSubcellularLocationCv [http]
cv:D006056 [http]	"Golgi Apparatus"^^xsd:string	:MeshAnatomyCv [http]
cv:GO_0044431 [http]	"Golgi apparatus part"^^xsd:string	:GoCellularComponentCv [http]

neXtProt – modifying the query

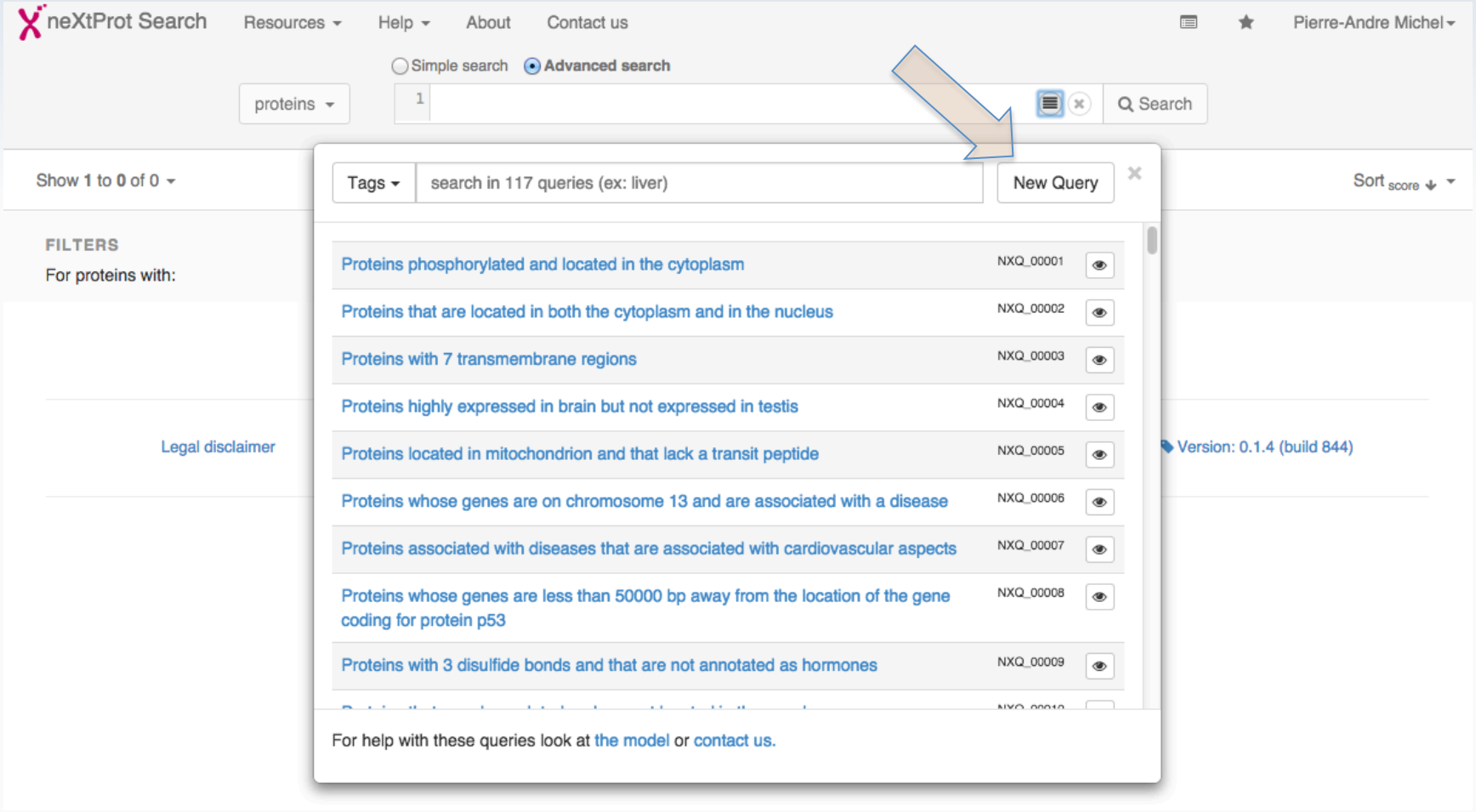


neXtProt – running the query

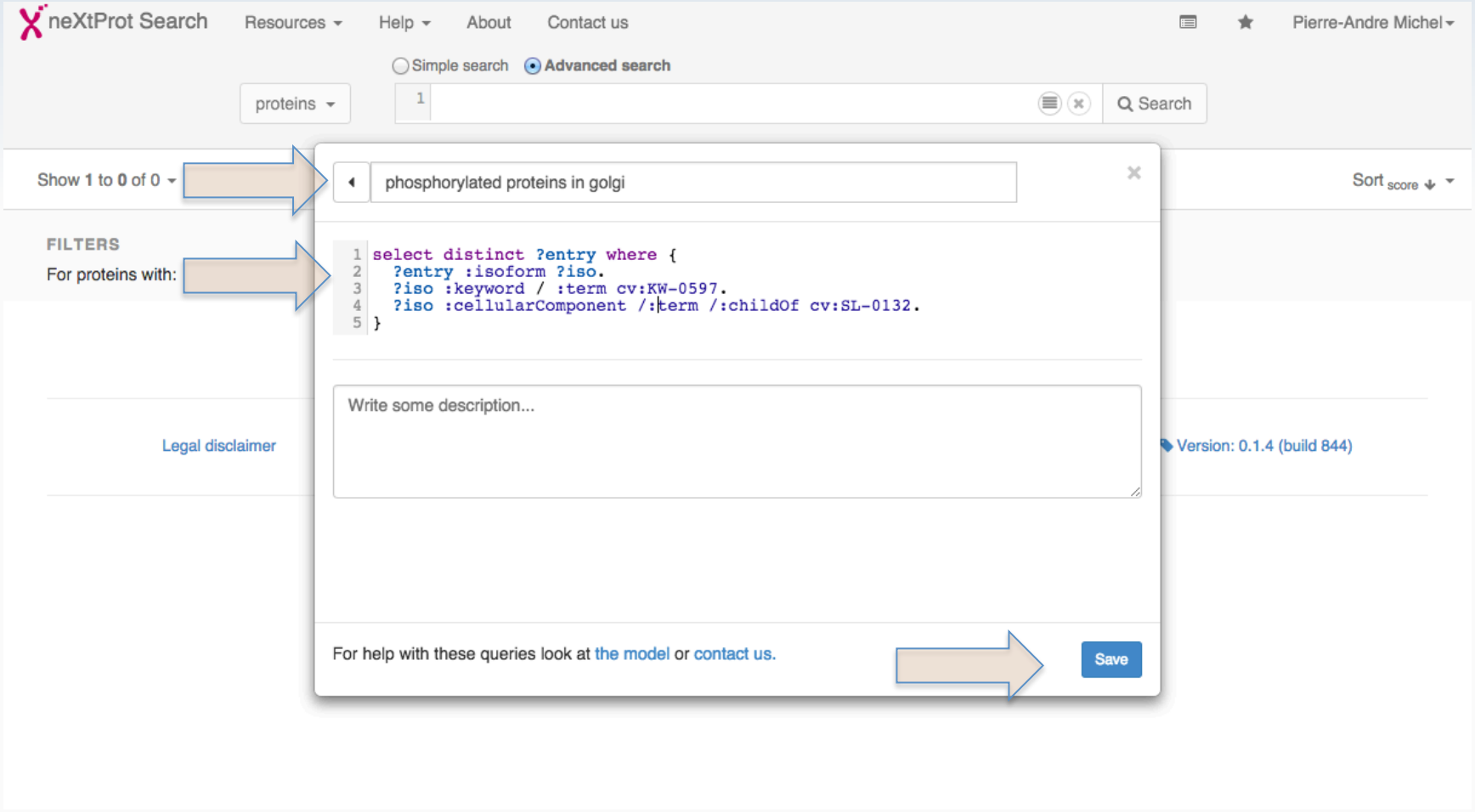
neXtProt – running the query



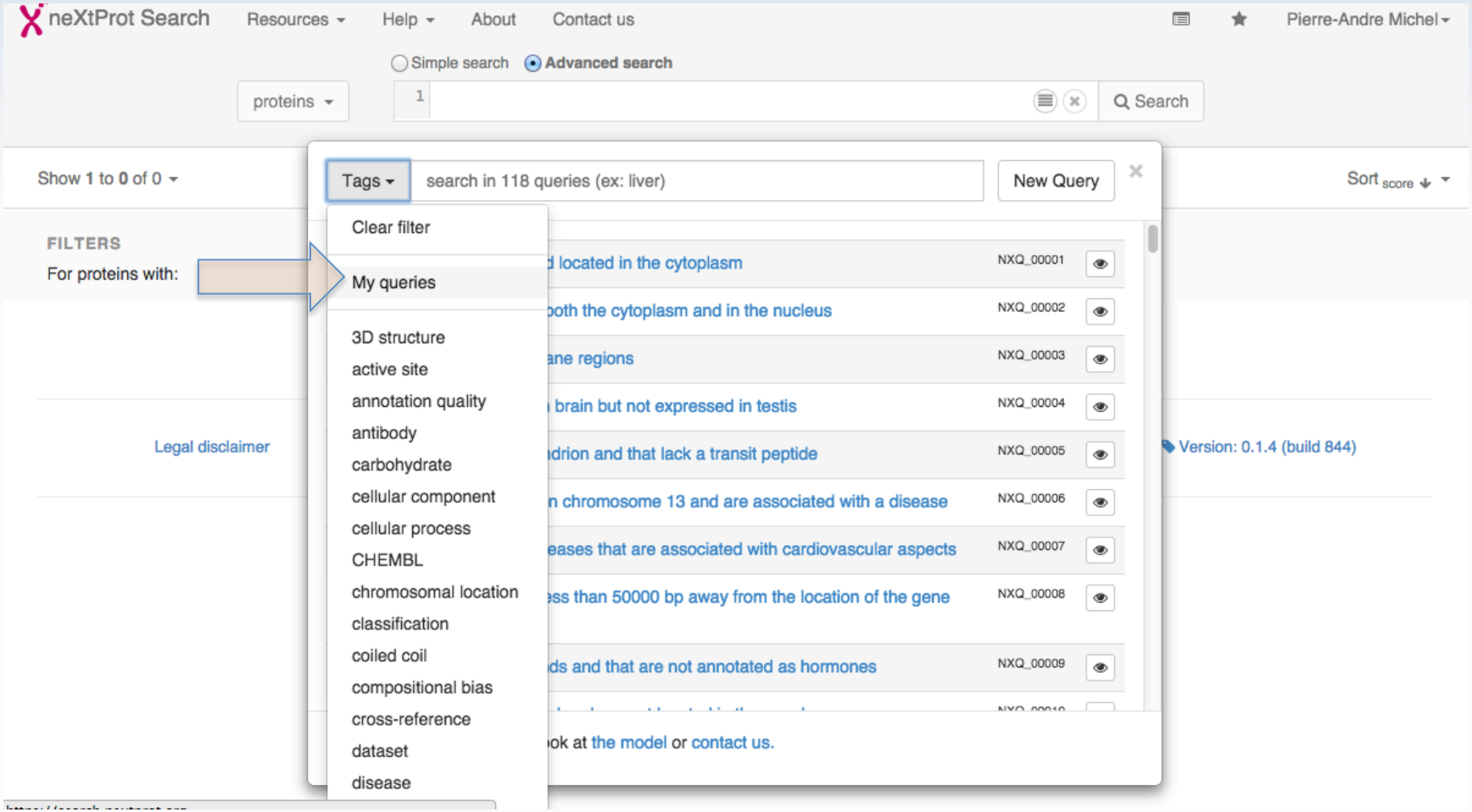
neXtProt – saving your query



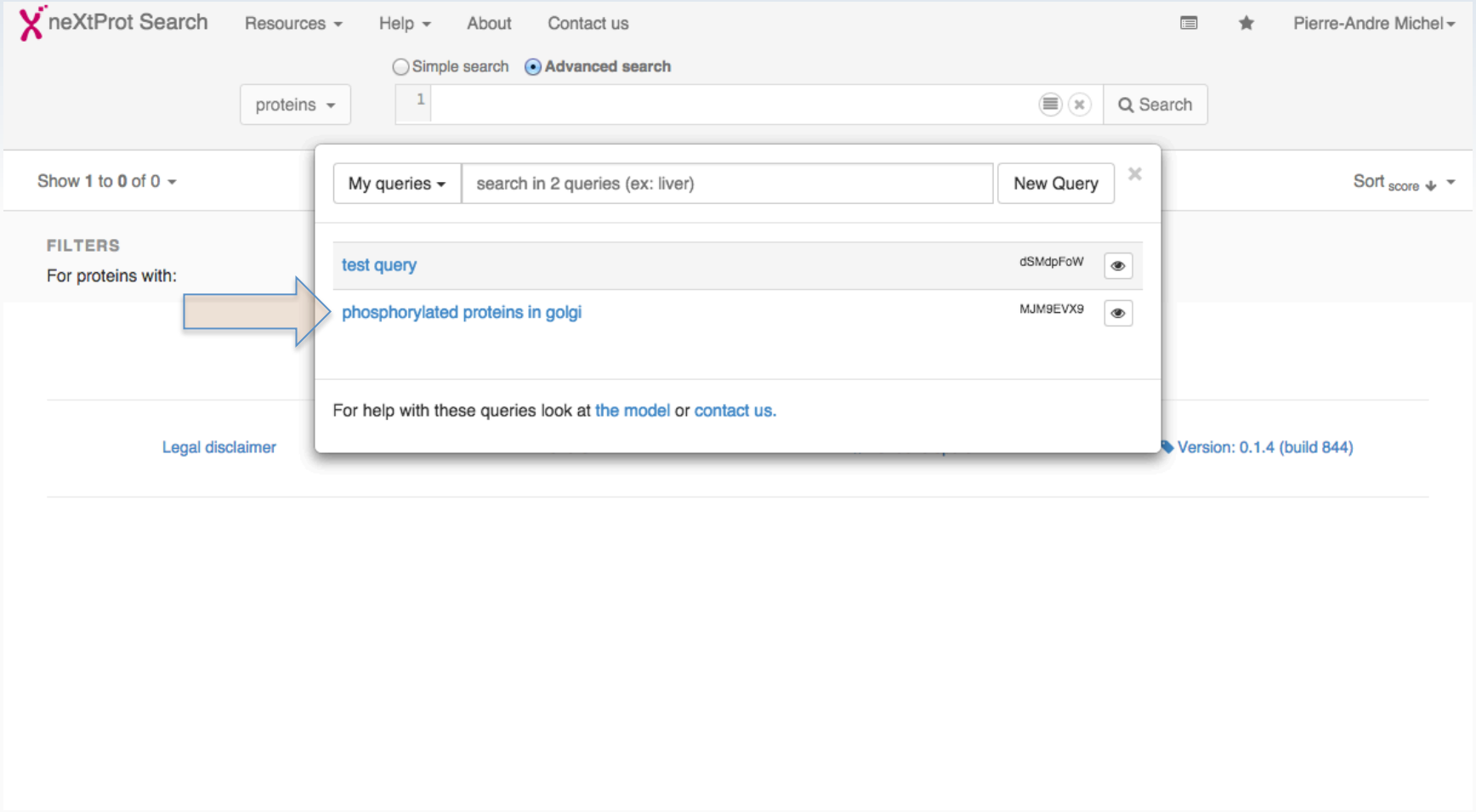
neXtProt – saving your query



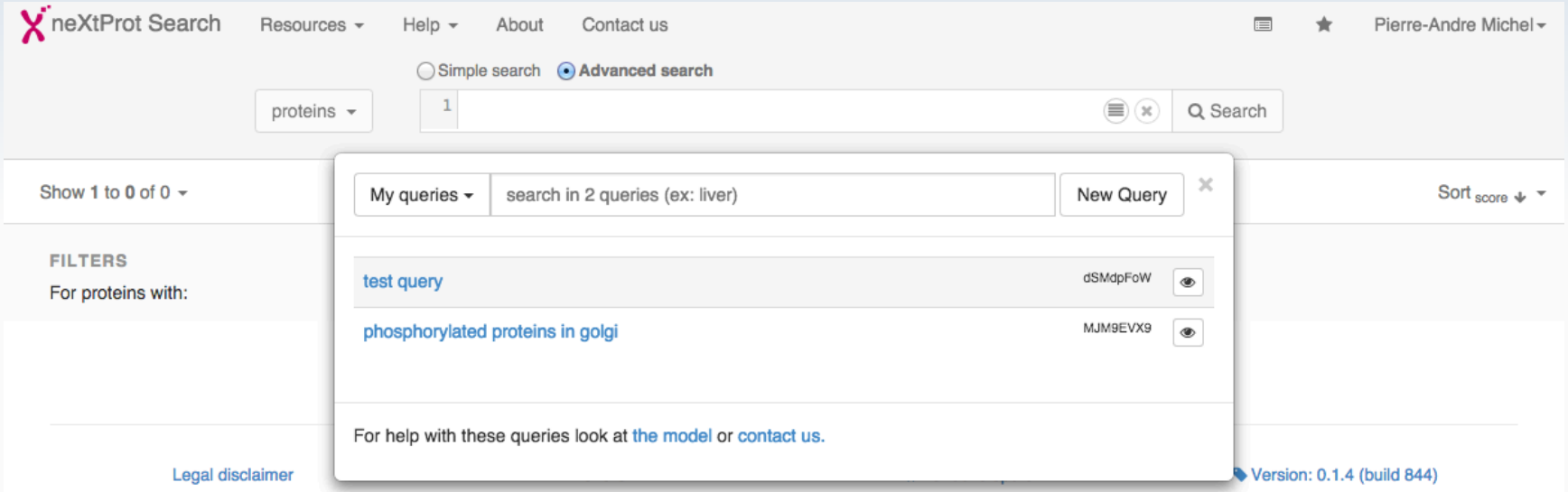
neXtProt – reusing your query



neXtProt – reusing your query



neXtProt – reusing your query

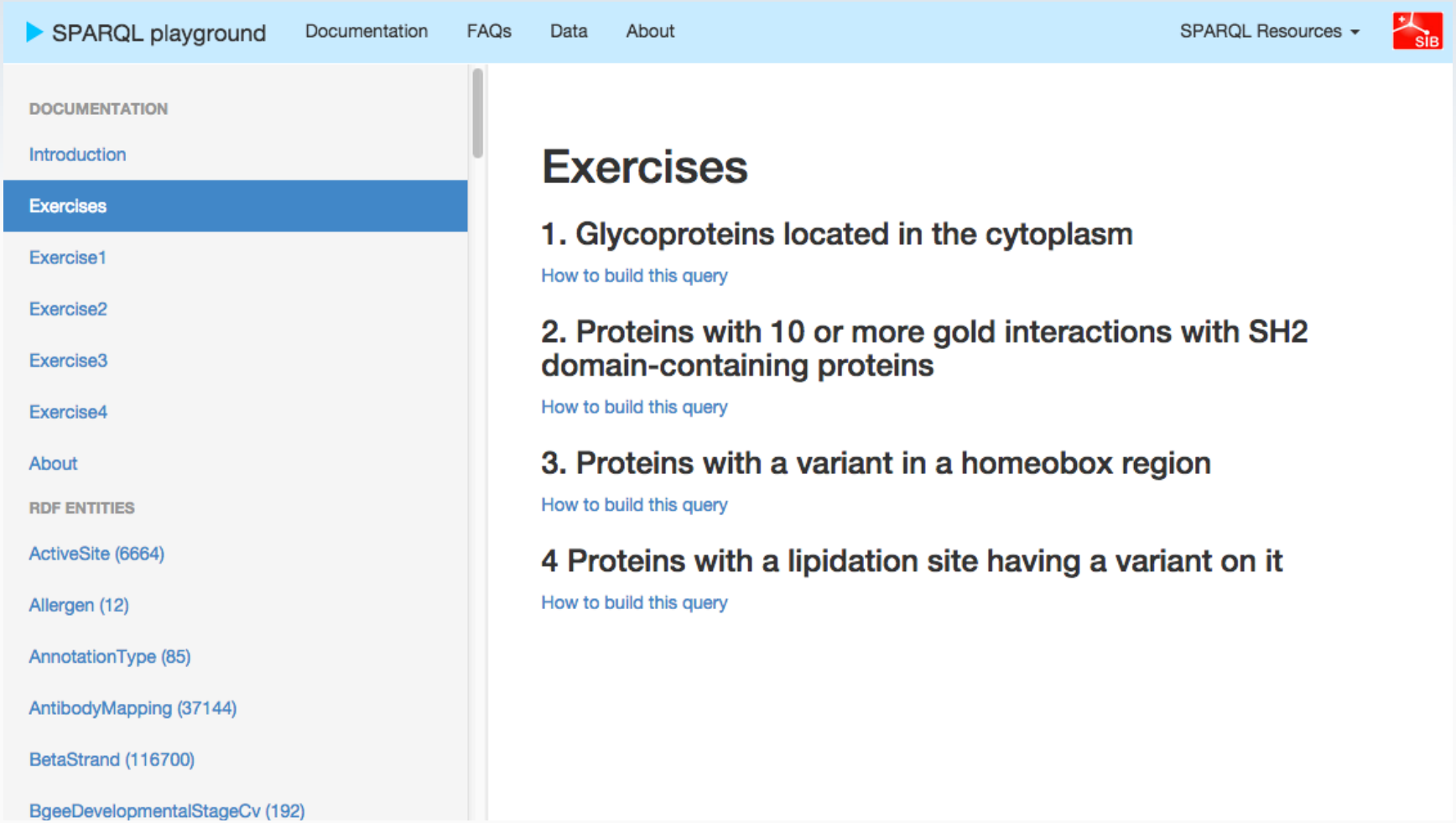


Sharable URL for your query

https://search.nextprot.org/proteins/search?mode=advanced&queryId=MJM9EVX9

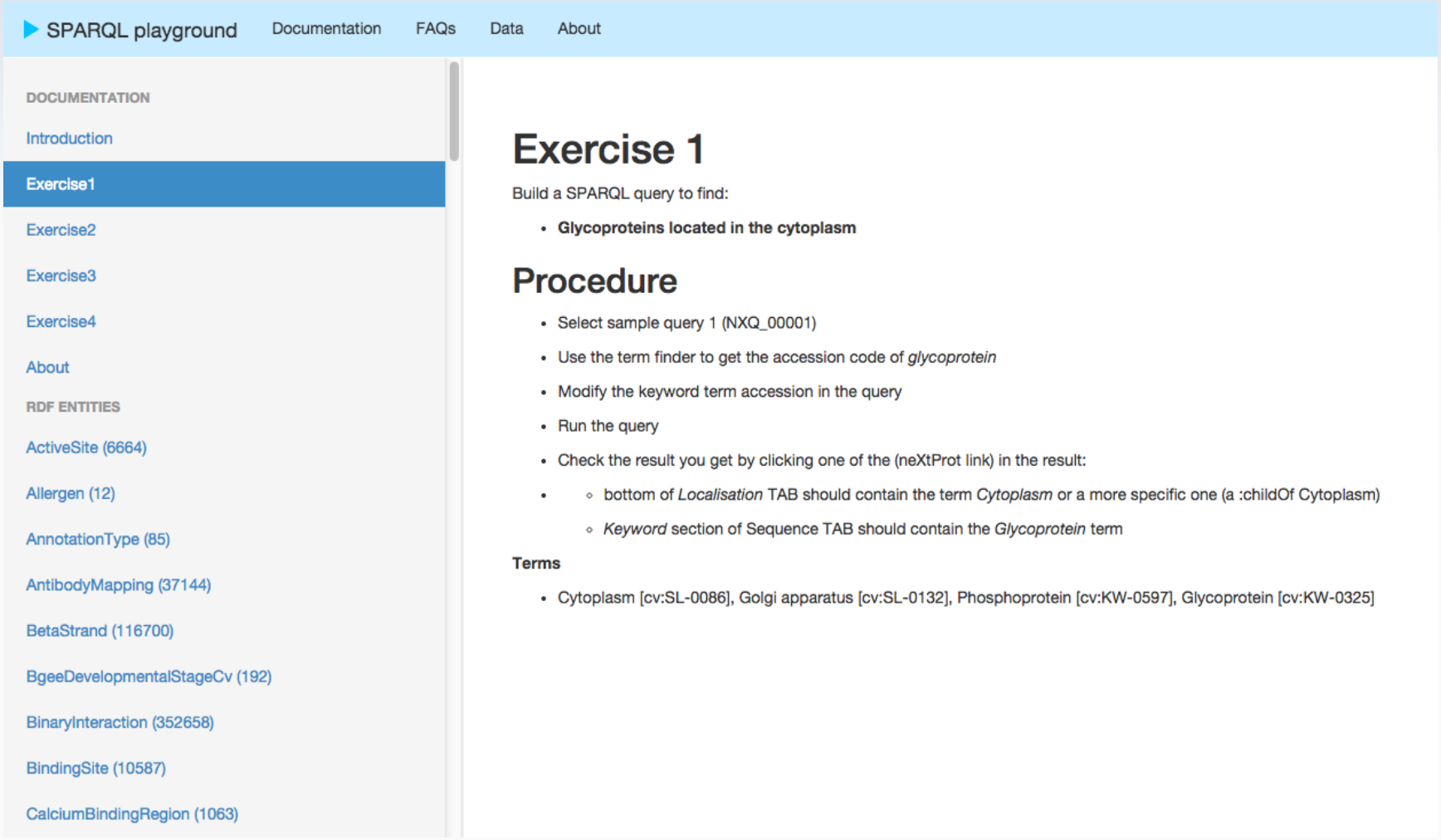
Now do it yourself !

http://sparql-playground.nextprot.org



Exercises

http://sparql-playground.nextprot.org/help/doc/exercise1



The end

Many thanks to

Jerven Bolleman

Daniel Teixeira

Alain Gateau

Monique Zahn

Pascale Gaudet