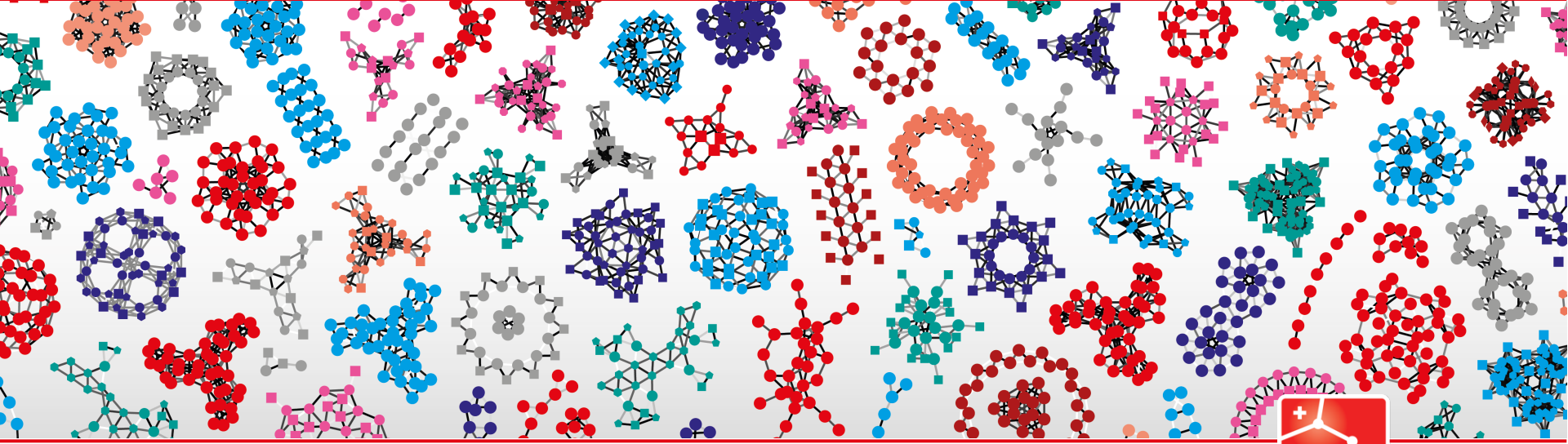


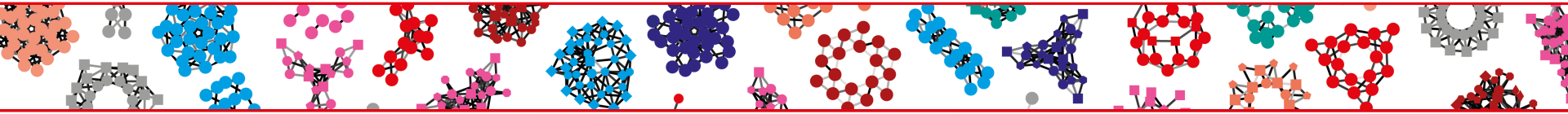
Swiss Institute of Bioinformatics

SIB → Elixir friends

Jerven Bolleman & Anne Morgat,

elixir SWITZERLAND

Overview

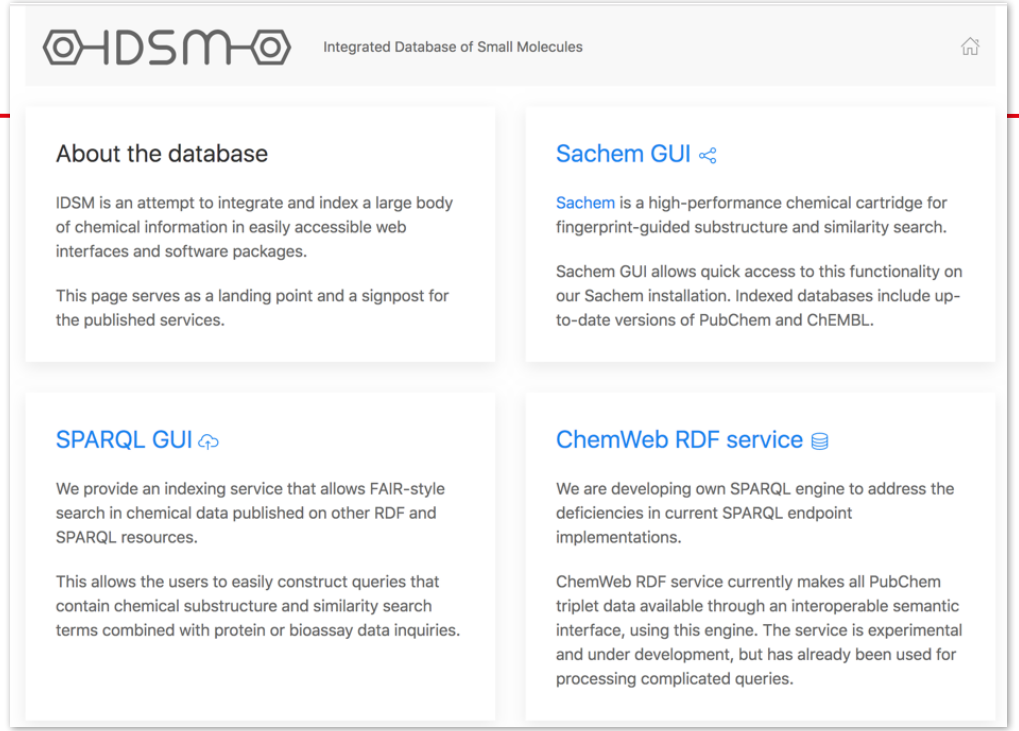


01 — **IDSM Elixir Czech node**

02 — **EBI RDF platform/Ensembl**

03 — **DisGeNET**

https://idsm.elixir-czech.cz/



Maintainers

Jakub Galgonek

Mirek Kratochvíl

Contact

jakub.galgonek@uochb.cas.cz

Bioinformatics group

IOCB CAS CZ, Prague

IDSM Service

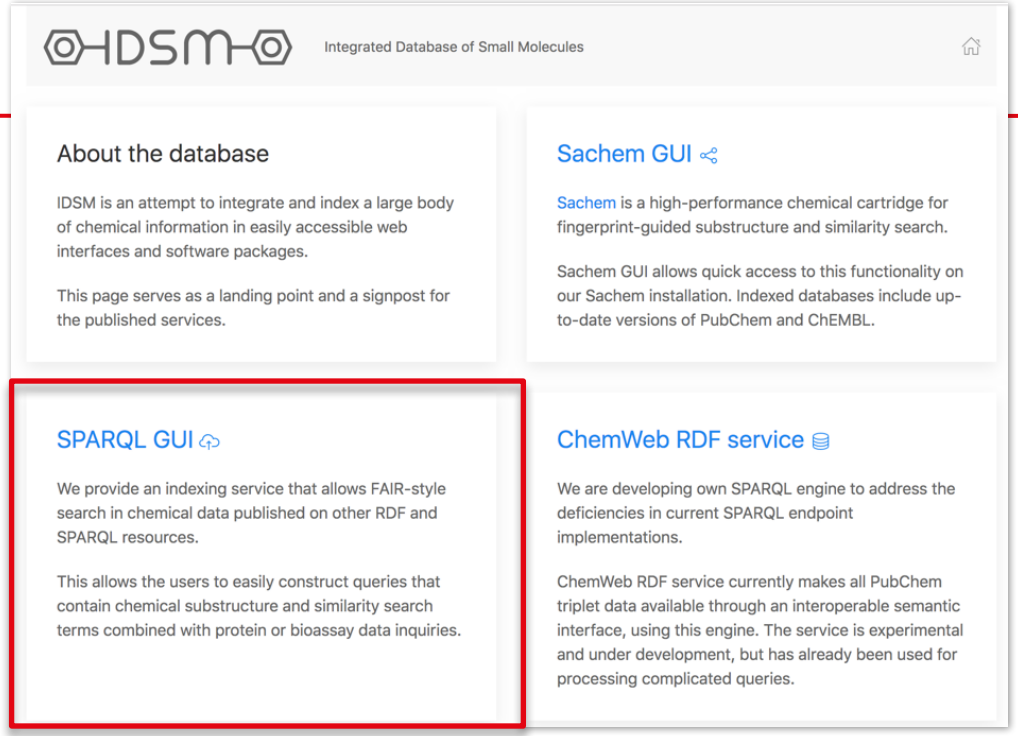
All services

Data privacy

ÚOCHB AV ČR
IOCB PRAGUE

elixir
CZECH REPUBLIC

https://idsm.elixir-czech.cz/



Maintainers

Jakub Galgonek

Mirek Kratochvíl

Contact

jakub.galgonek@uochb.cas.cz

Bioinformatics group

IOCB CAS CZ, Prague

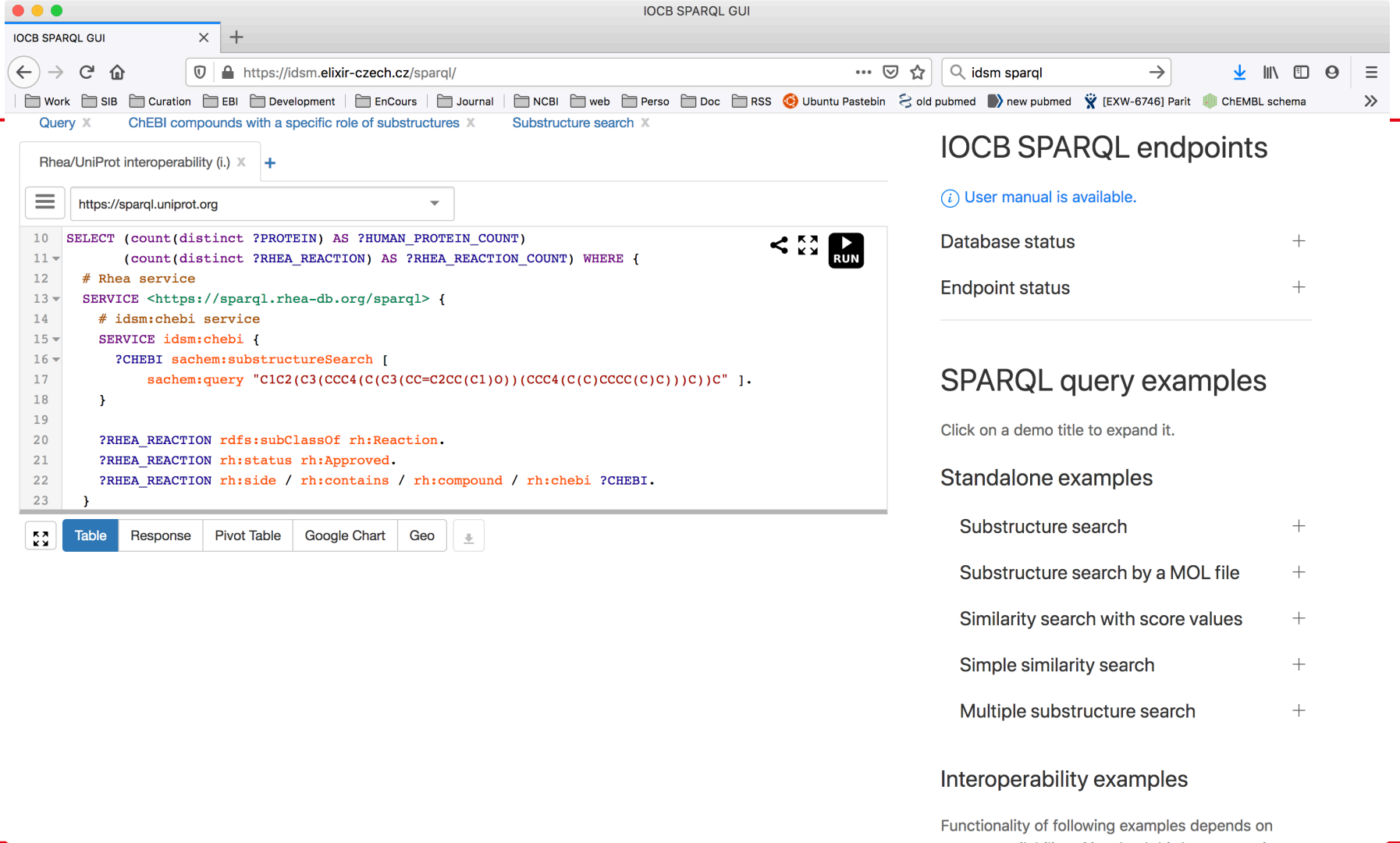
IDSM Service

All services

Data privacy

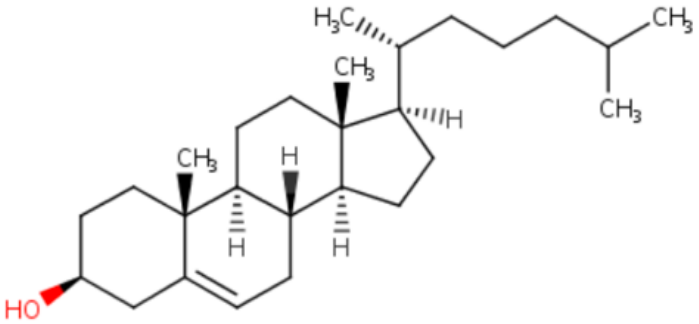
ÚOCHB AV ČR IOCB PRAGUE

elixir CZECH REPUBLIC


IOCB SPARQL GUI
IOCB SPARQL GUI
https://idsm.elixir-czech.cz/sparql/
idsm sparql
Work
SIB
Curation
EBI
Development
EnCours
Journal
NCBI
web
Perso
Doc
RSS
Ubuntu Pastebin
old pubmed
new pubmed
[EXW-6746] Parit
ChEMBL schema
Query
ChEBI compounds with a specific role of substructures
Substructure search
Rhea/UniProt interoperability (i.)
https://sparql.uniprot.org
SELECT (count(distinct ?PROTEIN) AS ?HUMAN_PROTEIN_COUNT)
(count(distinct ?RHEA_REACTION) AS ?RHEA_REACTION_COUNT) WHERE {
# Rhea service
SERVICE <https://sparql.rhea-db.org/sparql> {
# idsm:chebi service
SERVICE idsm:chebi {
?CHEBI sachem:substructureSearch [
sachem:query "C1C2(C3(CCC4(C(C3(CC=C2CC(C1)O))(CCC4(C(C)CCCC(C)C)))C))C"].
}
?RHEA_REACTION rdfs:subClassOf rh:Reaction.
?RHEA_REACTION rh:status rh:Approved.
?RHEA_REACTION rh:side / rh:contains / rh:compound / rh:chebi ?CHEBI.
}
RUN
Table
Response
Pivot Table
Google Chart
Geo
IOCB SPARQL endpoints
User manual is available.
Database status
Endpoint status
SPARQL query examples
Click on a demo title to expand it.
Standalone examples
Substructure search
Substructure search by a MOL file
Similarity search with score values
Simple similarity search
Multiple substructure search
Interoperability examples
Functionality of following examples depends on

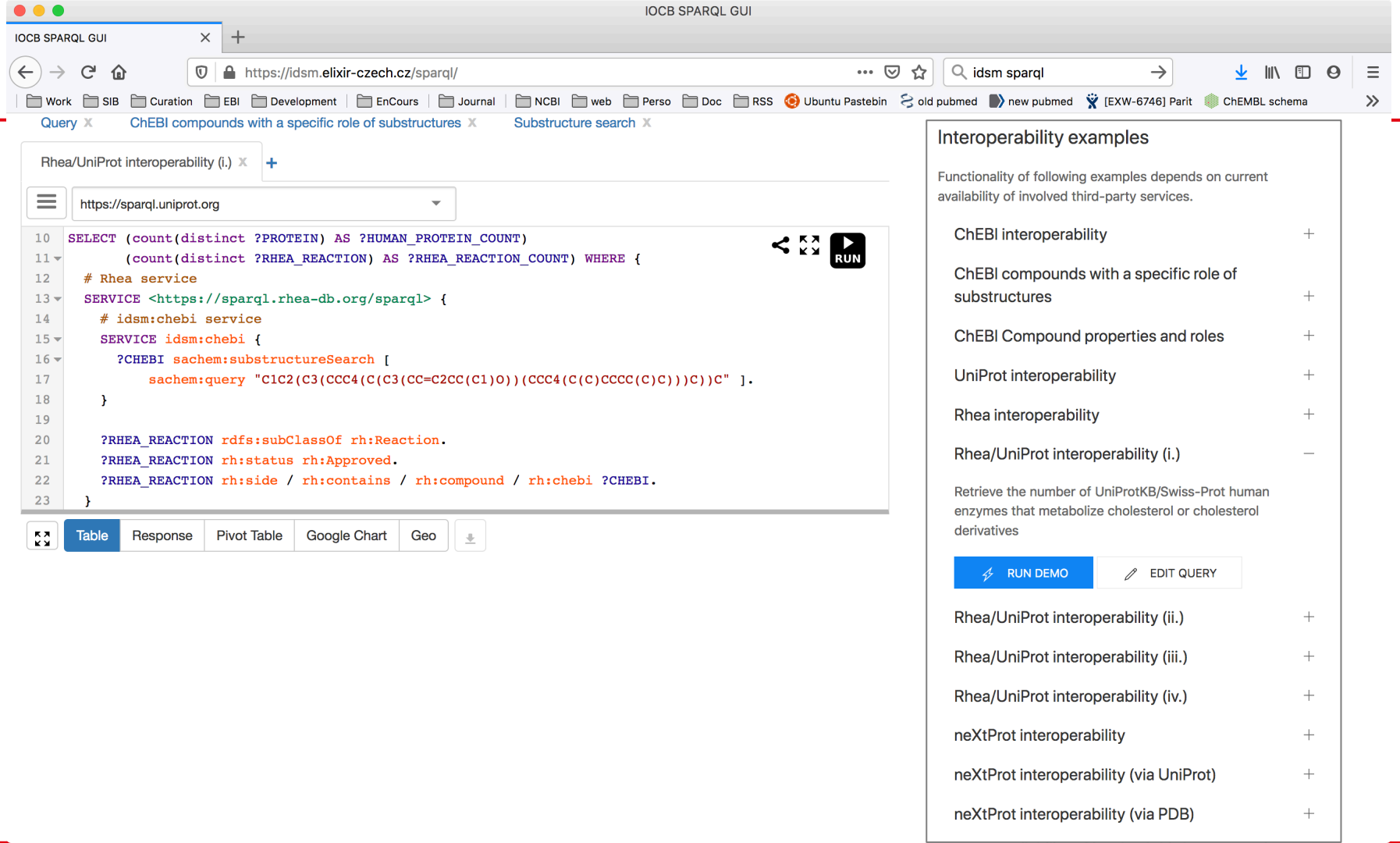
Chemical structure search using IDSM/Sachem

CHEBI:16113
(cholesterol)



Input: SMILES

C1[C@@]2([C@]3(CC[C@]4([C@]([C@@]3(CC=C2C[C@H](C1)O)[H])(CC[C@@]4([C@H](C)CCCC(C)C)[H])[H])C)[H])C


IOCB SPARQL GUI
IOCB SPARQL GUI
https://idsm.elixir-czech.cz/sparql/
idsm sparql
Work
SIB
Curation
EBI
Development
EnCours
Journal
NCBI
web
Perso
Doc
RSS
Ubuntu Pastebin
old pubmed
new pubmed
[EXW-6746] Parit
ChEMBL schema
Query
ChEBI compounds with a specific role of substructures
Substructure search
Rhea/UniProt interoperability (i.)
https://sparql.uniprot.org
SELECT (count(distinct ?PROTEIN) AS ?HUMAN_PROTEIN_COUNT)
(count(distinct ?RHEA_REACTION) AS ?RHEA_REACTION_COUNT) WHERE {
# Rhea service
SERVICE <https://sparql.rhea-db.org/sparql> {
# idsm:chebi service
SERVICE idsm:chebi {
?CHEBI sachem:substructureSearch [
sachem:query "C1C2(C3(CCC4(C(C3(CC=C2CC(C1)O))(CCC4(C(C)CCCC(C)C)))C))C"].
}
?RHEA_REACTION rdfs:subClassOf rh:Reaction.
?RHEA_REACTION rh:status rh:Approved.
?RHEA_REACTION rh:side / rh:contains / rh:compound / rh:chebi ?CHEBI.
}
RUN
Table
Response
Pivot Table
Google Chart
Geo
Interoperability examples
Functionality of following examples depends on current availability of involved third-party services.
ChEBI interoperability
ChEBI compounds with a specific role of substructures
ChEBI Compound properties and roles
UniProt interoperability
Rhea interoperability
Rhea/UniProt interoperability (i.)
Retrieve the number of UniProtKB/Swiss-Prot human enzymes that metabolize cholesterol or cholesterol derivatives
RUN DEMO
EDIT QUERY
Rhea/UniProt interoperability (ii.)
Rhea/UniProt interoperability (iii.)
Rhea/UniProt interoperability (iv.)
neXtProt interoperability
neXtProt interoperability (via UniProt)
neXtProt interoperability (via PDB)

Q30: Retrieve the Rhea reactions that involve cholesterol or cholesterol derivatives

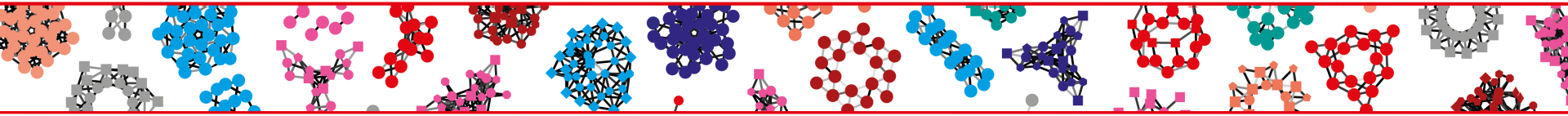
```
#endpoint:https://sparql.rhea-db.org/sparql

PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX sachem:<http://bioinfo.uochb.cas.cz/rdf/v1.0/sachem#>
PREFIX idsm:<https://idsm.elixir-czech.cz/sparql/endpoint/>
PREFIX up:<http://purl.uniprot.org/core/>
PREFIX rh:<http://rdf.rhea-db.org/>
SELECT ?chebi ?chebiUniprotName
            ?rhReaction ?rhReactionEquation
WHERE {
  SERVICE idsm:chebi {
    ?chebi sachem:substructureSearch
    [ sachem:query "C1[C@@]2([C@]3(CC[C@]4([C@]([C@@]3(CC=C2C[C@H](C1)O)[H])(CC[C@@]4([C@H](C)CCCC(C)C)[H])[H])C)[H])C" ] .
  }
  ?rhReaction rh:equation ?rhReactionEquation .
  ?rhReaction rh:status ?status .
  VALUES (?status) {(rh:Approved) (rh:Preliminary)}
  ?rhReaction rh:side ?reactionSide .
  ?reactionSide rh:contains ?participant .
  ?participant rh:compound ?compound .
  ?compound rh:chebi ?chebi .
  ?chebi up:name ?chebiUniprotName .
}
```

Q30: Retrieve the Rhea reactions that involve cholesterol or cholesterol derivatives

chebi	chebiUniprotName	rheaReaction	rheaReactionEquation
http://purl.obolibrary.org/obo/CHEBI_16113	"cholesterol"	http://rdf.rhea-db.org/58264	"a beta-D-glucosyl-(1<->1')-N-acylsphing-4-enine + cholesterol = an N-acylsphing-4-enine + cholesteryl-beta-D-gl
http://purl.obolibrary.org/obo/CHEBI_17495	"cholesteryl-beta-D-glucoside"	http://rdf.rhea-db.org/58264	"a beta-D-glucosyl-(1<->1')-N-acylsphing-4-enine + cholesterol = an N-acylsphing-4-enine + cholesteryl-beta-D-gl
http://purl.obolibrary.org/obo/CHEBI_16113	"cholesterol"	http://rdf.rhea-db.org/58316	"beta-D-glucosyl-N-hexadecanoylsphing-4E-enine + cholesterol = cholesteryl-beta-D-glucoside + N-hexadecanoylsphi
http://purl.obolibrary.org/obo/CHEBI_17495	"cholesteryl-beta-D-glucoside"	http://rdf.rhea-db.org/58316	"beta-D-glucosyl-N-hexadecanoylsphing-4E-enine + cholesterol = cholesteryl-beta-D-glucoside + N-hexadecanoylsphi
http://purl.obolibrary.org/obo/CHEBI_16113	"cholesterol"	http://rdf.rhea-db.org/58324	"beta-D-glucosyl-N-(9Z-octadecenoyl)-sphing-4E-enine + cholesterol = cholesteryl-beta-D-glucoside + N-(9Z-octade
http://purl.obolibrary.org/obo/CHEBI_17495	"cholesteryl-beta-D-glucoside"	http://rdf.rhea-db.org/58324	"beta-D-glucosyl-N-(9Z-octadecenoyl)-sphing-4E-enine + cholesterol = cholesteryl-beta-D-glucoside + N-(9Z-octade
http://purl.obolibrary.org/obo/CHEBI_16113	"cholesterol"	http://rdf.rhea-db.org/53468	"1-hexadecanoyl-2-(5Z,8Z,11Z,14Z,17Z-eicosapentaenoyl)-sn-glycero-3-phosphocholine + cholesterol = 1-hexadecanoy
http://purl.obolibrary.org/obo/CHEBI_84969	"cholesteryl (5Z,8Z,11Z,14Z,17Z-eicosapentaenoate)"	http://rdf.rhea-db.org/53468	"1-hexadecanoyl-2-(5Z,8Z,11Z,14Z,17Z-eicosapentaenoyl)-sn-glycero-3-phosphocholine + cholesterol = 1-hexadecanoy
http://purl.obolibrary.org/obo/CHEBI_16113	"cholesterol"	http://rdf.rhea-db.org/53472	"1-hexadecanoyl-2-(9Z,12Z-octadecadienoyl)-sn-glycero-3-phosphocholine + cholesterol = 1-hexadecanoyl-sn-glycero
http://purl.obolibrary.org/obo/CHEBI_41509	"cholesteryl (9Z,12Z)-octadecadienoate"	http://rdf.rhea-db.org/53472	"1-hexadecanoyl-2-(9Z,12Z-octadecadienoyl)-sn-glycero-3-phosphocholine + cholesterol = 1-hexadecanoyl-sn-glycero
http://purl.obolibrary.org/obo/CHEBI_16113	"cholesterol"	http://rdf.rhea-db.org/21104	"AH2 + cholesterol + O2 = 25-hydroxycholesterol + A + H2O"

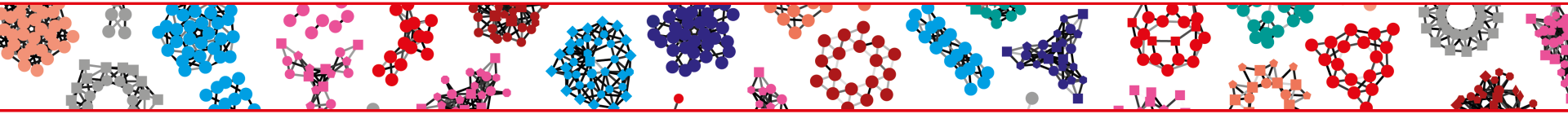
[…]



Accompanying Jupyter notebook:

https://github.com/sib-swiss/sparql-training/tree/master/rhea/SWAT4HCLS_2019

Overview



01 IDSM Elixir Czech node

02 EBI RDF platform/Ensembl

03 DisGeNET

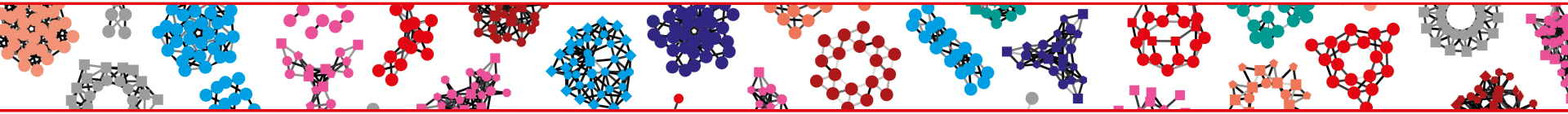
To Ensembl to get exon locations

https://sparql.uniprot.org/sparql/

```
PREFIX uniprotkb:<http://purl.uniprot.org/uniprot/>
PREFIX ensemblterms:<http://rdf.ebi.ac.uk/terms/ensembl/>
PREFIX obo:<http://purl.obolibrary.org/obo/>
PREFIX faldo:<http://biohackathon.org/resource/faldo#>

SELECT
   ?protein   ?transcript    ?begin    ?end
WHERE {
   BIND(uniprotkb:P05067 AS ?protein)
   SERVICE <https://www.ebi.ac.uk/rdf/services/sparql> {
      ?ensemblGene ensemblterms:DEPENDENT ?protein.
      ?transcript faldo:location ?location ;
                  obo:SO_transcribed_from ?ensemblGene .
      ?location faldo:begin [faldo:position ?begin] ;
                faldo:end [faldo:position ?end ] .
   }
}
```

Overview



01 IDSM Elixir Czech node

02 EBI RDF platform/Ensembl

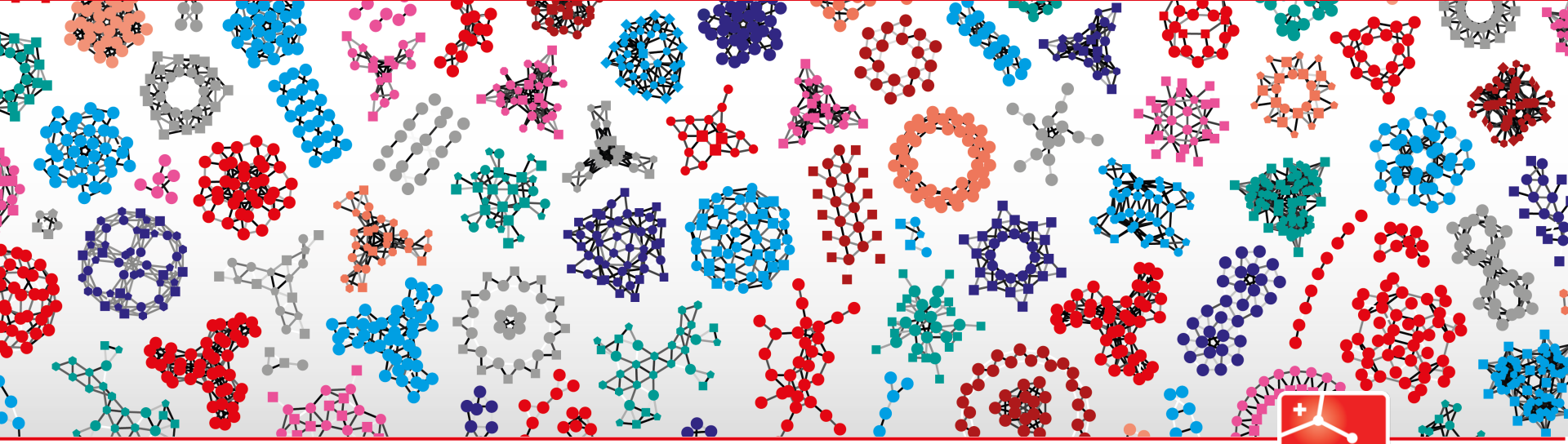
03 DisGeNET

Protein-Gene-Disease as in DisGeNET

https://sparql.uniprot.org/sparql/

```
PREFIX up:<http://purl.uniprot.org/core/>
PREFIX uniprotkb:<http://purl.uniprot.org/uniprot/>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX database:<http://purl.uniprot.org/database/>
PREFIX sio:<http://semanticscience.org/resource/>
PREFIX ncit:<http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#>
SELECT
   ?protein ?gene ?disease ?go
WHERE {
 BIND (uniprotkb:P31327 AS ?protein)
 ?protein a up:Protein ;
          up:classifiedWith ?go .
 ?go a owl:Class .

 SERVICE <http://rdf.disgenet.org/sparql/>{
    ?gda sio:SIO_000628 ?disease, ?gene .
?disease rdf:type ncit:C7057  .
   ?gene rdf:type ncit:C16612 ;
        sio:SIO_010078 ?protein .
}}
```



SIB
Swiss Institute of Bioinformatics

Thank you for your attention

elixir SWITZERLAND

www.sib.swiss

Sachem

2018

J Cheminform. 2018 May 23;10(1):27. doi: 10.1186/s13321-018-0282-y.

Sachem: a chemical cartridge for high-performance substructure search.

Kratochvíl M[1,2], Vondrášek J[1], Galgonek J[3].

2019

J Cheminform. 2019 Jun 28;11(1):45. doi: 10.1186/s13321-019-0367-2.

Interoperable chemical structure search service.

Kratochvíl M[1,2], Vondrášek J[1], Galgonek J[3].

Author information

1 Institute of Organic Chemistry and Biochemistry of the CAS, Flemingovo náměstí 2, 166 10, Prague 6, Czech Republic.

2 Department of Software Engineering, Faculty of Mathematics and Physics, Charles University, Malostranské náměstí 25, 118 00, Prague 1, Czech Republic.

3 Institute of Organic Chemistry and Biochemistry of the CAS, Flemingovo náměstí 2, 166 10, Prague 6, Czech Republic. galgonek@iocb.cas.cz.

Abstract

MOTIVATION: The existing connections between large databases of chemicals, proteins, metabolites and assays offer valuable resources for research in fields ranging from drug design to metabolomics. Transparent search across multiple databases provides a way to efficiently utilize these resources. To simplify such searches, many databases have adopted semantic technologies that allow interoperable querying of the datasets using SPARQL query language. However, the interoperable interfaces of the chemical databases still lack the functionality of structure-driven chemical search, which is a fundamental method of data discovery in the chemical search space.

RESULTS: We present a SPARQL service that augments existing semantic services by making interoperable substructure and similarity searches in small-molecule databases possible. The service thus offers new possibilities for querying interoperable databases, and simplifies writing of heterogeneous queries that include chemical-structure search terms.

AVAILABILITY: The service is freely available and accessible using a standard SPARQL endpoint interface. The service documentation and user-oriented demonstration interfaces that allow quick explorative querying of datasets are available at https://idsm.elixir-czech.cz .