SPARQL Query Formulation through BioFed Interface

Availability: http://vmurq09.deri.ie:8007/

With the Standard interface, the user can formulate basic SPARQL query. Whereas the interface can process simple as well as complex federated queries over Life Science Linked Open Data Cloud (LS-LOD).

Example 1: A biologist interested to query geneName, Name and Protein Sequence for enzymes

```
SELECT distinct * WHERE
 {
?x a <http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugbank/enzymes> .
?x <http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugbank/geneName> ?y .
?x <http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugbank/name> ?z .
?x <http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugbank/proteinSequence> ?n .
}
limit 10
```

In order to make such query the user first selects the concept “enzymes” from the “Make Selection” drop box (Figure 1).

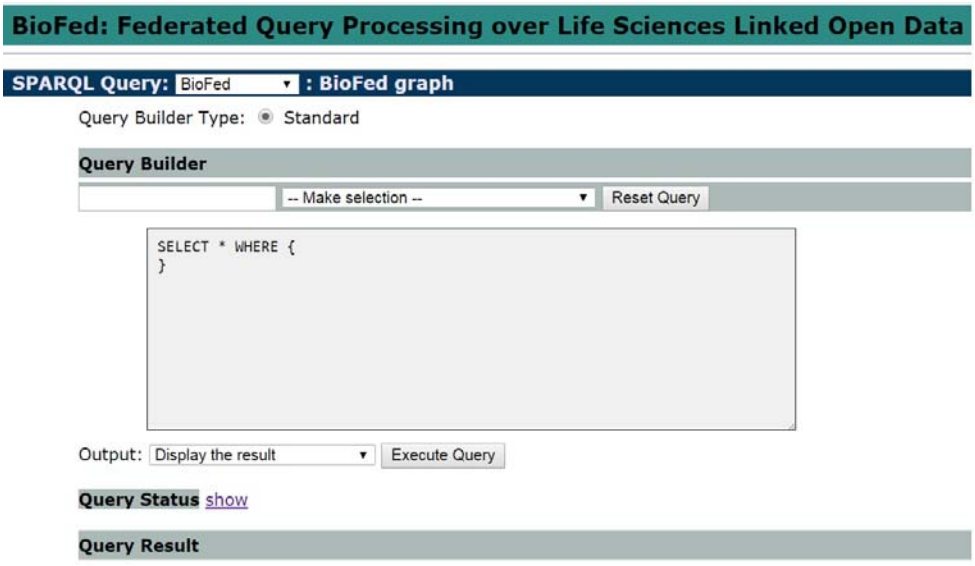


Figure 1: BioFed main interface

After selecting the concept “enzymes”, he selects “geneName”, “Name” and “Protein Sequence” properties from the “Make Selection” drop box that will now show various properties associated

with the concept “enzymes” and add them in query text area through “Add to Query” button. User has to change/chose different variable names himself e.g in our example we choose ?**y, ?z, ?n.** All must be different (Figure 2).

```
SELECT distinct * WHERE {
?x a <http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugbank/enzymes> .
?x <http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugbank/geneName> ?y .
?x <http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugbank/name> ?z .
?x <http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugbank/proteinSequence> ?n .
}
limit 10
```

Figure 2: SPARQL query generation: "give me geneName, Name and Protein Sequence for enzymes " from BioFed interface

After making the query, the user can add “distinct”, “Filter”, “Limit” and “Order By” clause by himself. In our example we use “distinct” and “limit” our answers by 10 (Figure 3). By clicking on “Execute Query”, the results will be displayed (Figure 3).

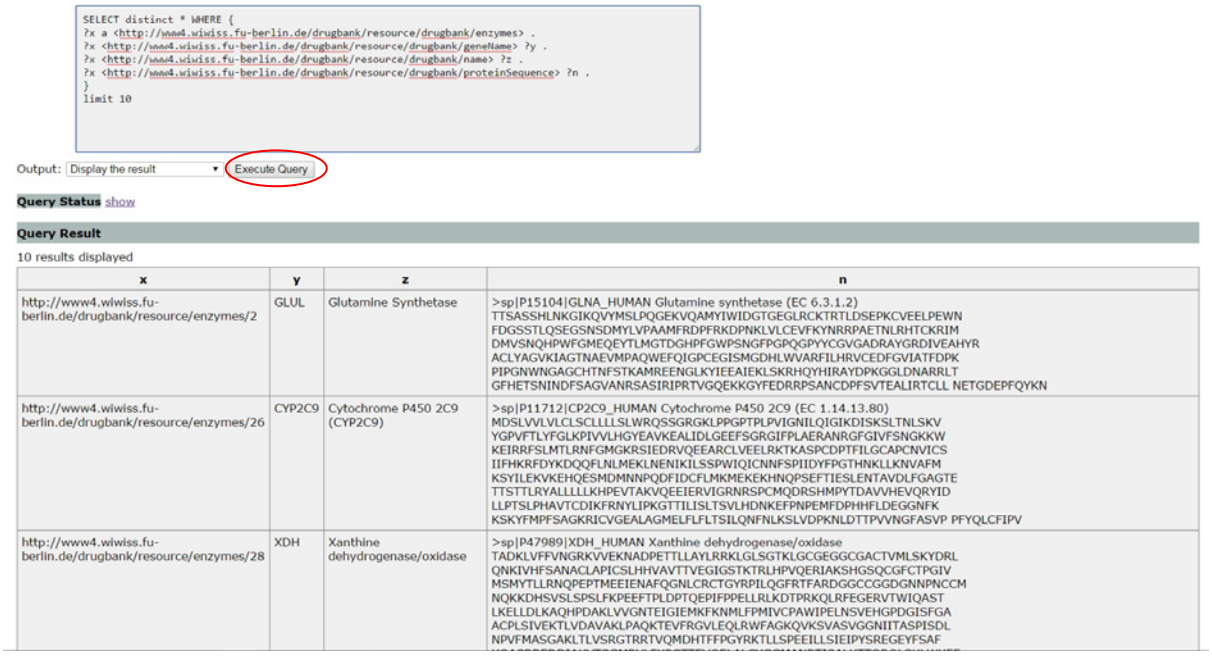


Figure 3: Results display through BioFed

The “Output” drop box contains two options “Display the results” and “Download the results”.

After selecting “Display the results” option, the user will click “Execute Query” button to display the results (Figure 3).

The user can also execute and download the query results in various formats such as CSV, JSON, XML, Text, TSV and Turtle (Figure 4 and Figure 5).

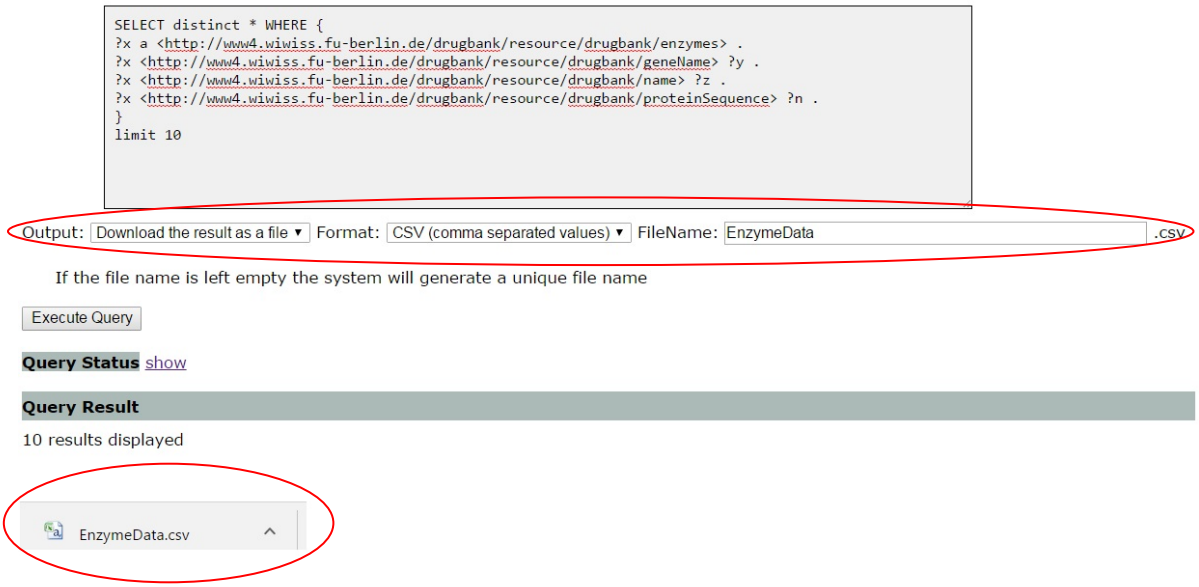


Figure 4: Results display in different formats

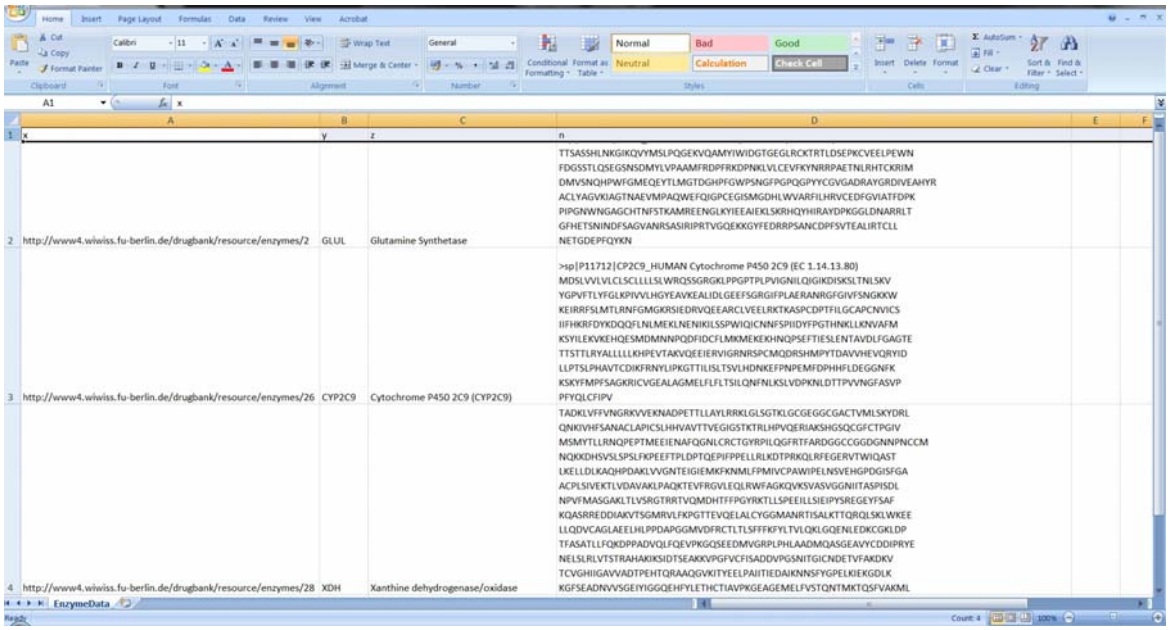


Figure 5: Downloaded results in CSV format

Example 2: Simple and Complex SPARQL query execution through BioFed

Instead of formulating the basic query from drop down menu if an expert user wants to execute any other query he can run such queries as well.

Different Simple and Complex queries are presented in the paper and listed at http://srvgal78.deri.ie/BioFed/queries.html

Let us consider Simple Query 2 and copy it to the interface and execute the query

SQL2: Find all the drug description and chemical equations of reactions related to drugs from category Cathartics.

```
PREFIX drugcategory: <http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugcategory/>
PREFIX drugbank: <http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugbank/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX kegg: <http://bio2rdf.org/ns/kegg#>
SELECT ?drugDesc ?cpd ?equation WHERE {
?drug drugbank:drugCategory drugcategory:cathartics .
?drug drugbank:keggCompoundId ?cpd .
?drug drugbank:description ?drugDesc .
?enzyme kegg:xSubstrate ?cpd .
?enzyme  rdf:type kegg:Enzyme .
?reaction kegg:xEnzyme ?enzyme .
?reaction kegg:equation ?equation . }
```

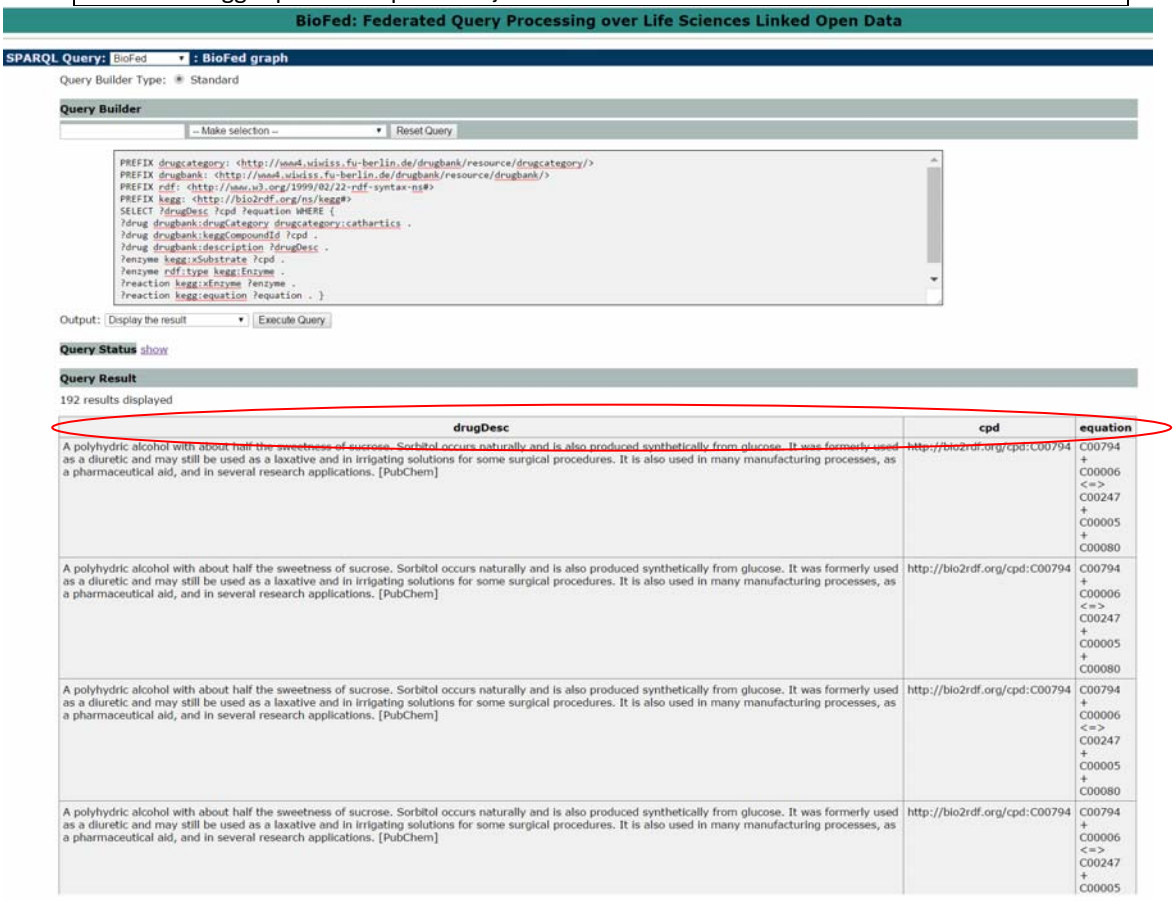
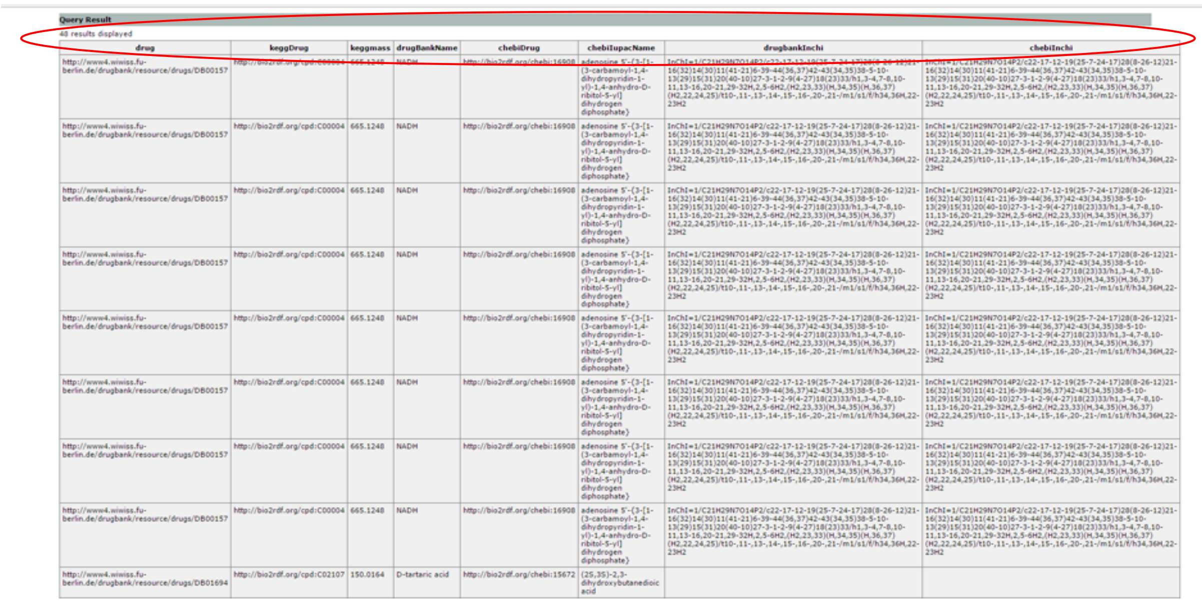


Figure 6: Simple Federated SPARQL query execution through BioFed Interface

Let us consider Complex Query 2 and copy it to the interface and execute the query

CQ2: Find all the drugs with their mass and chebi IupacName optionally the Inchi values retrieving from two sources are equal.

```
PREFIX drugbank: <http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugbank/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX chebi: <http://bio2rdf.org/ns/chebi#>
PREFIX purl: <http://purl.org/dc/elements/1.1/>
PREFIX bio2RDF: <http://bio2rdf.org/ns/bio2rdf#>
SELECT * WHERE  {
?drug rdf:type drugbank:drugs .
?drug drugbank:keggCompoundId ?keggDrug .
?keggDrug bio2RDF:mass ?keggmass .
?drug drugbank:genericName ?drugBankName .
?chebiDrug purl:title ?drugBankName .
?chebiDrug chebi:iupacName ?chebiIupacName .
OPTIONAL {
?drug drugbank:inchiIdentifier ?drugbankInchi .
?chebiDrug bio2RDF:inchi ?chebiInchi.
FILTER (?drugbankInchi = ?chebiInchi) } }
```



Query Result

48 results displayed

drug	keggDrug	keggmass	drugBankName	chebiDrug	chebiIupacName	drugbankInchi	chebiInchi
http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugs/DB00157	http://bio2rdf.org/cpd:C00004	665.1248	NADH	http://bio2rdf.org/chebi:16908	adenosine 5'-{3-[1-(3-carbamoyl-1,4-dihydropyridin-1-yl)-1,4-anhydro-D-ribitol-5-yl] dihydrogen diphosphate}	InChI=1/C21H29N7O14P2/c22-17-12-19(25-7-24-17)28(8-26-12)21-16(32)14(30)11(41-21)6-39-44(36,37)42-43(34,35)38-5-10-13(29)15(31)20(40-10)27-3-1-2-9(4-27)18(23)33/h1,3-4,7-8,10-11,13-16,20-21,29-32H,2,5-6H2,(H2,23,33)(H,34,35)(H,36,37)(H2,22,24,25)/t10-,11-,13-,14-,15-,16-,20-,21-/m1/s1/f/h34,36H,22-23H2	InChI=1/C21H29N7O14P2/c22-17-12-19(25-7-24-17)28(8-26-12)21-16(32)14(30)11(41-21)6-39-44(36,37)42-43(34,35)38-5-10-13(29)15(31)20(40-10)27-3-1-2-9(4-27)18(23)33/h1,3-4,7-8,10-11,13-16,20-21,29-32H,2,5-6H2,(H2,23,33)(H,34,35)(H,36,37)(H2,22,24,25)/t10-,11-,13-,14-,15-,16-,20-,21-/m1/s1/f/h34,36H,22-23H2
http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugs/DB00157	http://bio2rdf.org/cpd:C00004	665.1248	NADH	http://bio2rdf.org/chebi:16908	adenosine 5'-{3-[1-(3-carbamoyl-1,4-dihydropyridin-1-yl)-1,4-anhydro-D-ribitol-5-yl] dihydrogen diphosphate}	InChI=1/C21H29N7O14P2/c22-17-12-19(25-7-24-17)28(8-26-12)21-16(32)14(30)11(41-21)6-39-44(36,37)42-43(34,35)38-5-10-13(29)15(31)20(40-10)27-3-1-2-9(4-27)18(23)33/h1,3-4,7-8,10-11,13-16,20-21,29-32H,2,5-6H2,(H2,23,33)(H,34,35)(H,36,37)(H2,22,24,25)/t10-,11-,13-,14-,15-,16-,20-,21-/m1/s1/f/h34,36H,22-23H2	InChI=1/C21H29N7O14P2/c22-17-12-19(25-7-24-17)28(8-26-12)21-16(32)14(30)11(41-21)6-39-44(36,37)42-43(34,35)38-5-10-13(29)15(31)20(40-10)27-3-1-2-9(4-27)18(23)33/h1,3-4,7-8,10-11,13-16,20-21,29-32H,2,5-6H2,(H2,23,33)(H,34,35)(H,36,37)(H2,22,24,25)/t10-,11-,13-,14-,15-,16-,20-,21-/m1/s1/f/h34,36H,22-23H2
http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugs/DB00157	http://bio2rdf.org/cpd:C00004	665.1248	NADH	http://bio2rdf.org/chebi:16908	adenosine 5'-{3-[1-(3-carbamoyl-1,4-dihydropyridin-1-yl)-1,4-anhydro-D-ribitol-5-yl] dihydrogen diphosphate}	InChI=1/C21H29N7O14P2/c22-17-12-19(25-7-24-17)28(8-26-12)21-16(32)14(30)11(41-21)6-39-44(36,37)42-43(34,35)38-5-10-13(29)15(31)20(40-10)27-3-1-2-9(4-27)18(23)33/h1,3-4,7-8,10-11,13-16,20-21,29-32H,2,5-6H2,(H2,23,33)(H,34,35)(H,36,37)(H2,22,24,25)/t10-,11-,13-,14-,15-,16-,20-,21-/m1/s1/f/h34,36H,22-23H2	InChI=1/C21H29N7O14P2/c22-17-12-19(25-7-24-17)28(8-26-12)21-16(32)14(30)11(41-21)6-39-44(36,37)42-43(34,35)38-5-10-13(29)15(31)20(40-10)27-3-1-2-9(4-27)18(23)33/h1,3-4,7-8,10-11,13-16,20-21,29-32H,2,5-6H2,(H2,23,33)(H,34,35)(H,36,37)(H2,22,24,25)/t10-,11-,13-,14-,15-,16-,20-,21-/m1/s1/f/h34,36H,22-23H2
http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugs/DB00157	http://bio2rdf.org/cpd:C00004	665.1248	NADH	http://bio2rdf.org/chebi:16908	adenosine 5'-{3-[1-(3-carbamoyl-1,4-dihydropyridin-1-yl)-1,4-anhydro-D-ribitol-5-yl] dihydrogen diphosphate}	InChI=1/C21H29N7O14P2/c22-17-12-19(25-7-24-17)28(8-26-12)21-16(32)14(30)11(41-21)6-39-44(36,37)42-43(34,35)38-5-10-13(29)15(31)20(40-10)27-3-1-2-9(4-27)18(23)33/h1,3-4,7-8,10-11,13-16,20-21,29-32H,2,5-6H2,(H2,23,33)(H,34,35)(H,36,37)(H2,22,24,25)/t10-,11-,13-,14-,15-,16-,20-,21-/m1/s1/f/h34,36H,22-23H2	InChI=1/C21H29N7O14P2/c22-17-12-19(25-7-24-17)28(8-26-12)21-16(32)14(30)11(41-21)6-39-44(36,37)42-43(34,35)38-5-10-13(29)15(31)20(40-10)27-3-1-2-9(4-27)18(23)33/h1,3-4,7-8,10-11,13-16,20-21,29-32H,2,5-6H2,(H2,23,33)(H,34,35)(H,36,37)(H2,22,24,25)/t10-,11-,13-,14-,15-,16-,20-,21-/m1/s1/f/h34,36H,22-23H2
http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugs/DB00157	http://bio2rdf.org/cpd:C00004	665.1248	NADH	http://bio2rdf.org/chebi:16908	adenosine 5'-{3-[1-(3-carbamoyl-1,4-dihydropyridin-1-yl)-1,4-anhydro-D-ribitol-5-yl] dihydrogen diphosphate}	InChI=1/C21H29N7O14P2/c22-17-12-19(25-7-24-17)28(8-26-12)21-16(32)14(30)11(41-21)6-39-44(36,37)42-43(34,35)38-5-10-13(29)15(31)20(40-10)27-3-1-2-9(4-27)18(23)33/h1,3-4,7-8,10-11,13-16,20-21,29-32H,2,5-6H2,(H2,23,33)(H,34,35)(H,36,37)(H2,22,24,25)/t10-,11-,13-,14-,15-,16-,20-,21-/m1/s1/f/h34,36H,22-23H2	InChI=1/C21H29N7O14P2/c22-17-12-19(25-7-24-17)28(8-26-12)21-16(32)14(30)11(41-21)6-39-44(36,37)42-43(34,35)38-5-10-13(29)15(31)20(40-10)27-3-1-2-9(4-27)18(23)33/h1,3-4,7-8,10-11,13-16,20-21,29-32H,2,5-6H2,(H2,23,33)(H,34,35)(H,36,37)(H2,22,24,25)/t10-,11-,13-,14-,15-,16-,20-,21-/m1/s1/f/h34,36H,22-23H2
http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugs/DB00157	http://bio2rdf.org/cpd:C00004	665.1248	NADH	http://bio2rdf.org/chebi:16908	adenosine 5'-{3-[1-(3-carbamoyl-1,4-dihydropyridin-1-yl)-1,4-anhydro-D-ribitol-5-yl] dihydrogen diphosphate}	InChI=1/C21H29N7O14P2/c22-17-12-19(25-7-24-17)28(8-26-12)21-16(32)14(30)11(41-21)6-39-44(36,37)42-43(34,35)38-5-10-13(29)15(31)20(40-10)27-3-1-2-9(4-27)18(23)33/h1,3-4,7-8,10-11,13-16,20-21,29-32H,2,5-6H2,(H2,23,33)(H,34,35)(H,36,37)(H2,22,24,25)/t10-,11-,13-,14-,15-,16-,20-,21-/m1/s1/f/h34,36H,22-23H2	InChI=1/C21H29N7O14P2/c22-17-12-19(25-7-24-17)28(8-26-12)21-16(32)14(30)11(41-21)6-39-44(36,37)42-43(34,35)38-5-10-13(29)15(31)20(40-10)27-3-1-2-9(4-27)18(23)33/h1,3-4,7-8,10-11,13-16,20-21,29-32H,2,5-6H2,(H2,23,33)(H,34,35)(H,36,37)(H2,22,24,25)/t10-,11-,13-,14-,15-,16-,20-,21-/m1/s1/f/h34,36H,22-23H2
http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugs/DB00157	http://bio2rdf.org/cpd:C00004	665.1248	NADH	http://bio2rdf.org/chebi:16908	adenosine 5'-{3-[1-(3-carbamoyl-1,4-dihydropyridin-1-yl)-1,4-anhydro-D-ribitol-5-yl] dihydrogen diphosphate}	InChI=1/C21H29N7O14P2/c22-17-12-19(25-7-24-17)28(8-26-12)21-16(32)14(30)11(41-21)6-39-44(36,37)42-43(34,35)38-5-10-13(29)15(31)20(40-10)27-3-1-2-9(4-27)18(23)33/h1,3-4,7-8,10-11,13-16,20-21,29-32H,2,5-6H2,(H2,23,33)(H,34,35)(H,36,37)(H2,22,24,25)/t10-,11-,13-,14-,15-,16-,20-,21-/m1/s1/f/h34,36H,22-23H2	InChI=1/C21H29N7O14P2/c22-17-12-19(25-7-24-17)28(8-26-12)21-16(32)14(30)11(41-21)6-39-44(36,37)42-43(34,35)38-5-10-13(29)15(31)20(40-10)27-3-1-2-9(4-27)18(23)33/h1,3-4,7-8,10-11,13-16,20-21,29-32H,2,5-6H2,(H2,23,33)(H,34,35)(H,36,37)(H2,22,24,25)/t10-,11-,13-,14-,15-,16-,20-,21-/m1/s1/f/h34,36H,22-23H2
http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugs/DB00157	http://bio2rdf.org/cpd:C00004	665.1248	NADH	http://bio2rdf.org/chebi:16908	adenosine 5'-{3-[1-(3-carbamoyl-1,4-dihydropyridin-1-yl)-1,4-anhydro-D-ribitol-5-yl] dihydrogen diphosphate}	InChI=1/C21H29N7O14P2/c22-17-12-19(25-7-24-17)28(8-26-12)21-16(32)14(30)11(41-21)6-39-44(36,37)42-43(34,35)38-5-10-13(29)15(31)20(40-10)27-3-1-2-9(4-27)18(23)33/h1,3-4,7-8,10-11,13-16,20-21,29-32H,2,5-6H2,(H2,23,33)(H,34,35)(H,36,37)(H2,22,24,25)/t10-,11-,13-,14-,15-,16-,20-,21-/m1/s1/f/h34,36H,22-23H2	InChI=1/C21H29N7O14P2/c22-17-12-19(25-7-24-17)28(8-26-12)21-16(32)14(30)11(41-21)6-39-44(36,37)42-43(34,35)38-5-10-13(29)15(31)20(40-10)27-3-1-2-9(4-27)18(23)33/h1,3-4,7-8,10-11,13-16,20-21,29-32H,2,5-6H2,(H2,23,33)(H,34,35)(H,36,37)(H2,22,24,25)/t10-,11-,13-,14-,15-,16-,20-,21-/m1/s1/f/h34,36H,22-23H2
http://www4.wiwiss.fu-berlin.de/drugbank/resource/drugs/DB01694	http://bio2rdf.org/cpd:C02107	150.0164	D-tartaric acid	http://bio2rdf.org/chebi:15672	(2S,3S)-2,3-dihydroxybutanedioic acid		

Figure 7: Complex Federated SPARQL query execution through BioFed Interface

Important Note

BioFed is designed to query real world life science publicly available SPARQL endpoints. The example previously mentioned in the paper was just an example and results are not guaranteed considering the remote endpoints are not under the control of one who query those but the publishers. This means if any of the SPARQL endpoint contributing to the result of any of the triple patterns mentioned in the query is not available the overall result cannot be constituted hence we will not get any answer.