

# SPARQL examples: SIB model, software, and patches

Egon Willighagen 

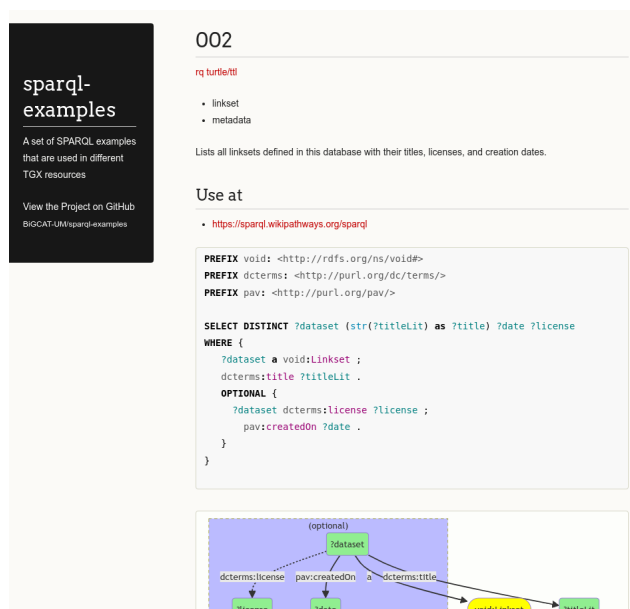
Published November 17, 2024

## Citation

Willighagen, E. (2024). SPARQL examples: SIB model, software, and patches. *Chem-bla-ics*. <https://doi.org/10.59350/djm89-5nb39>

## Keywords

Sparql, Wikipathways, Vhp4safety, ChEMBL, Scholia



The screenshot shows a web page for a SPARQL query titled '002'. On the left is a dark sidebar with the text 'sparql-examples', 'A set of SPARQL examples that are used in different TGX resources', and 'View the Project on GitHub: BIGCAT-UM/sparql-examples'. The main content area has a light yellow background. At the top, it says '002' and 'rq turtle/ttl'. Below that are two bullet points: 'linkset' and 'metadata'. A line of text reads 'Lists all linksets defined in this database with their titles, licenses, and creation dates.' Under the heading 'Use at', there is a link to 'https://sparql.wikipathways.org/sparql'. A code block contains the following SPARQL query: 

```
PREFIX void: <http://rdfs.org/ns/void#>
PREFIX dcterm: <http://purl.org/dc/terms/>
PREFIX pav: <http://purl.org/pav/>

SELECT DISTINCT ?dataset (str(?titleLit) as ?title) ?date ?license
WHERE {
  ?dataset a void:Linkset ;
  dcterm:title ?titleLit .
  OPTIONAL {
    ?dataset dcterm:license ?license ;
    pav:createdOn ?date .
  }
}
```

 Below the code is a diagram showing the query structure. A central node '?dataset' is connected to '?titleLit' via 'dcterm:title', to '?license' via 'dcterm:license', and to '?date' via 'pav:createdOn'. A dashed box labeled '(optional)' encloses the connections to '?license' and '?date'. A yellow node 'void:Linkset' is connected to '?dataset' via 'a', and to '?titleLit' via 'void:Linkset'.

## Copyright

Copyright © Egon Willighagen 2024. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## chem-bla-ics

Jerven Bolleman *et al.* recently [published a great preprint](#) about how to use RDF to give SPARQL queries context by linking it (semantically) with metadata. The context includes keywords, the SPARQL endpoint the query can be run against, and a human-oriented description of the query. A few groups have at recent hackathons been working on using the combination of a SPARQL query and a human-oriented description to train large language models, including the group behind this paper. Given that SPARQL is a very small language, I can see this may work well, and that it may support our [VHP4Safety](#) and [Scholia](#) projects.

But in addition to the data model for SPARQL as research output (see [doi:10.32388/ZNWI7T.2](https://doi.org/10.32388/ZNWI7T.2)), the paper also introduces the [sparql-example-utils](#) software that I was first introduced with at [the recent October Scholia hackathon](#).

But I have/had some features I like to see added. The first is provenance. Who is the author/contributor of the SPARQL query? Is there a open license for it, or perhaps public domain? How do I give attribution if I reuse the SPARQL query? These things matter in a modern [recognition and rewards](#) world where is room for everyone's talent. A set of good SPARQL queries may be more valuable than a ten-page Jupyter notebook (and the other way around). So, I [started writing patches](#). And I created [a custom jar](#) so that I can see these patches in action in [our growing list of SPARQL queries](#) (here [a WikiPathways query](#)):

I started collecting SPARQL queries for [ChEMBL](#), [WikiPathways](#), and [VHP4Safety](#). These queries are often part of other interfaces but we can easily extract the original SPARQL from the Turtle files behind these pages.