

Proto-OKN: Cross-Knowledge Graph Queries

Natural Language Queries using MCP Servers

Peter Rose (pwrose.ucsd@gmail.com)

SPOKE TEAM

PI: Sergio Baranzini (UC San Francisco)

Peter Rose (UCSD), Sui Huang (ISB), Scooter Morris (UCSF),
Angela Rizk-Jackson (UCSF), Yongmei Shi (UCSF), Sam Gebre (NASA),
Amanda Saravia-Butler (NASA), Kirill Grigorev (NASA),
Aenor Sawyer (UCSF/NASA), Charlotte Nelson (Mate Bioservices)

WOBD TEAM (The Scripps Research Institute, TSRI)

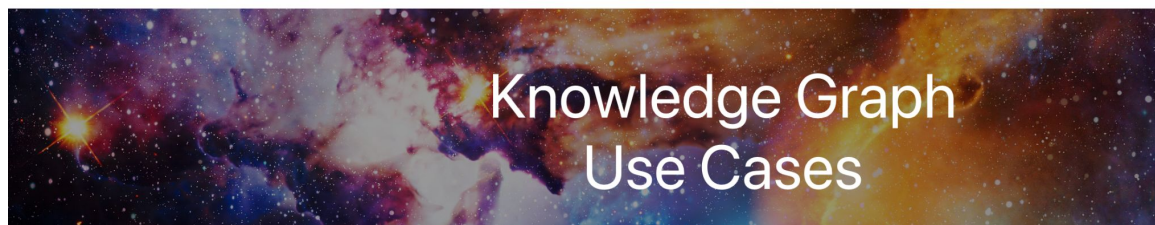
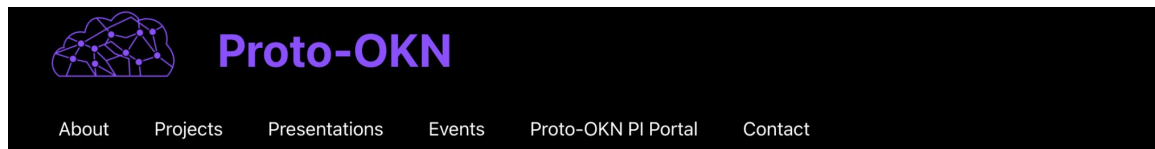
Andrew Su, Benjamin Good

FRINK TEAM (Renaissance Computing Institute, RENCi)

Christopher Bizon, Jim Balhoff, Yaphet Kebede

NSF Awards 2333819, 2535091

NSF Proto-OKN Program



“Open knowledge network (OKN) is a publicly accessible, interconnected set of data repositories and associated knowledge graphs that will enable data-driven, artificial intelligence-based solutions”

<https://www.proto-okn.net/projects/>

FRINK: SPARQL Query of the Proto-OKN KGs

FRINK
Query the Proto-OKN

Home

About

QUERY

Sources

FRINK Federated SPARQL

Custom Sources

SPARQL Query

```
1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 PREFIX schema: <https://purl.org/okn/frink/kg/spoke-okn/schema/>
3 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
4
5 SELECT DISTINCT ?location_label ?npl_status (COUNT(DISTINCT ?
compound) as ?compound_count)
6 FROM <https://purl.org/okn/frink/kg/spoke-okn>
7 WHERE {
8   ?stmt rdf:subject ?compound ;
9         rdf:predicate schema:FOUNDIN_CfL ;
10        rdf:object ?location ;
11        schema:metrics ?metrics ;
12        schema:value ?npl_status .
13
14   ?location rdfs:label ?location_label .
15
16   FILTER(CONTAINS(?metrics, "NPL Status"))
17   FILTER(CONTAINS(?npl_status, "NPL"))
18 }
19 GROUP BY ?location_label ?npl_status
20 ORDER BY DESC(?compound_count)
21 LIMIT 50
```

Save Query

Run Query

RESULTS

50 results in 55.24s

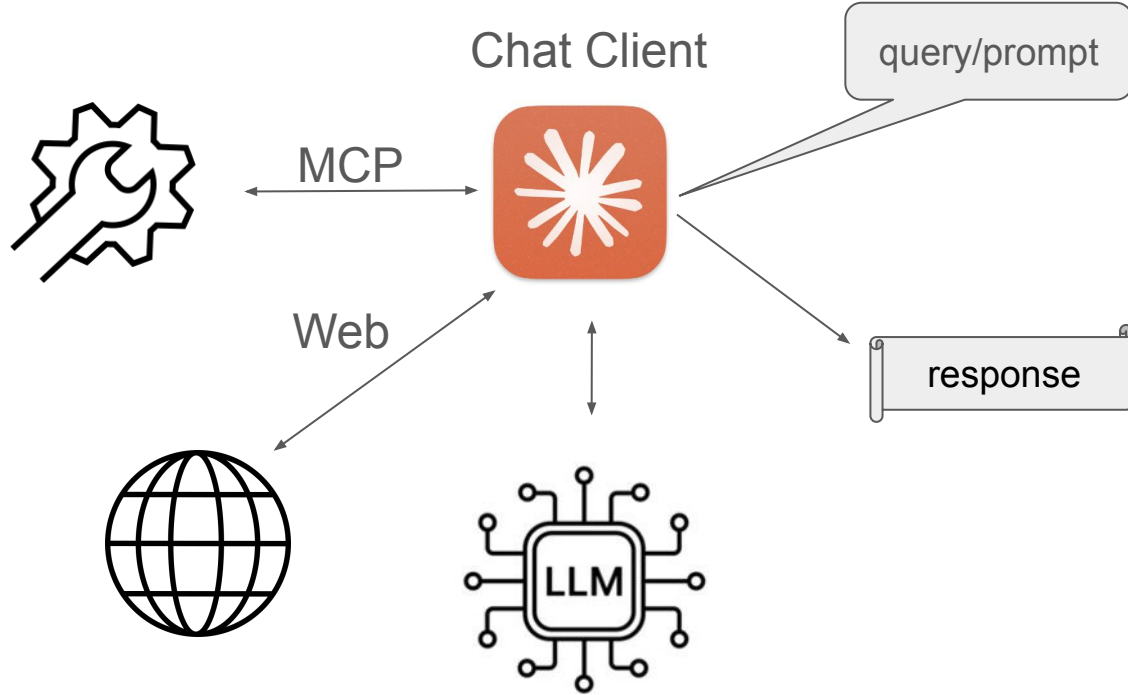
location_label	npl_status	compound_count
08805	Currently on the Final...	74
93429	Currently on the Final...	67
94124	Currently on the Final...	66
95376	Currently on the Final...	65
08641	Currently on the Final...	61
46514	Currently on the Final...	60
36108	Currently on the Final...	60
70510	Deleted from the Fina...	60
97231	Currently on the Final...	57
95670	Currently on the Final...	57
02542	Currently on the Final...	56
45502	Not on the NPL;Not o...	56
21010	Currently on the Final...	55
12901	Currently on the Final...	54

SPARQL queries are verbose and complex

Few users have SPARQL experience

<https://frink.apps.renci.org/>

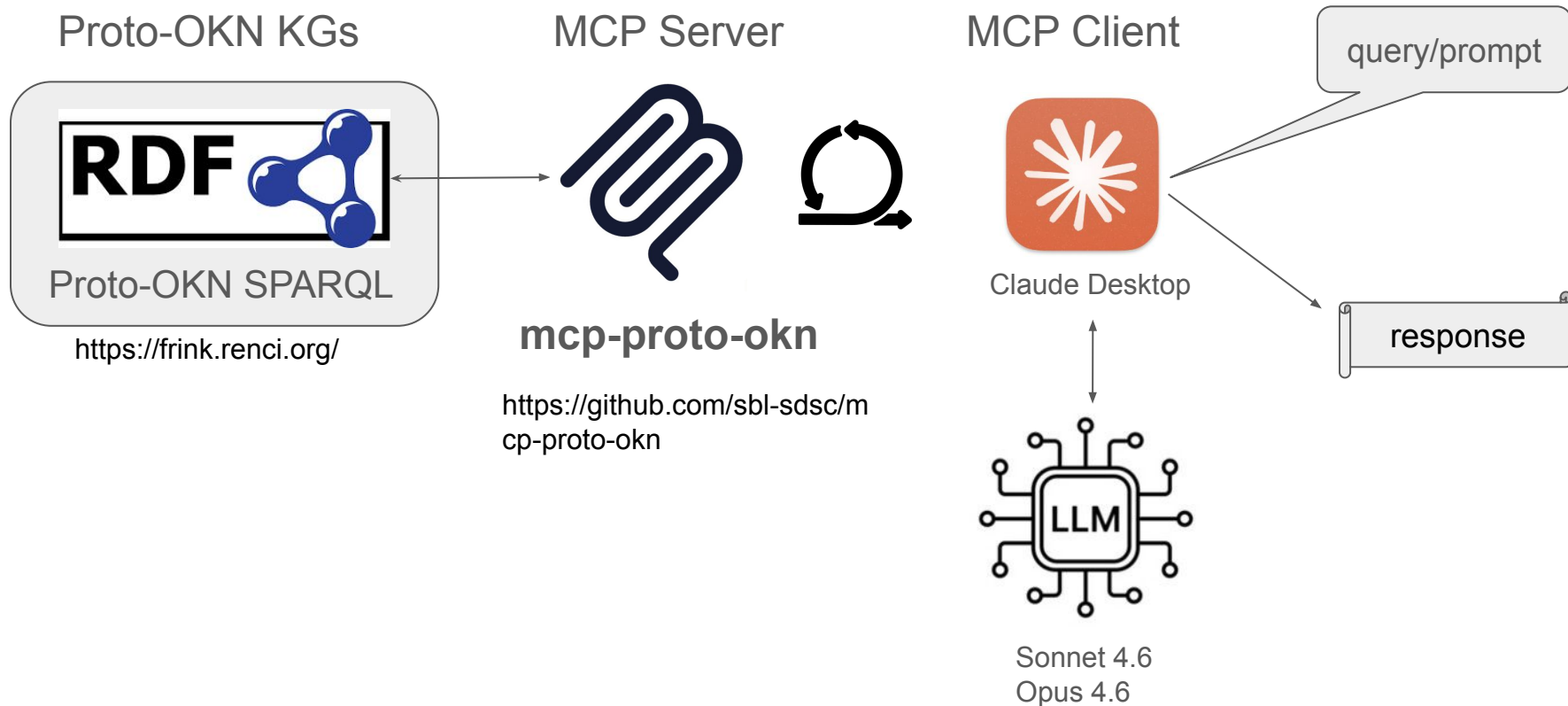
AI Assistants and Tools



MCP (Model Context Protocol) is an open protocol that standardizes how applications provide context to large language models.

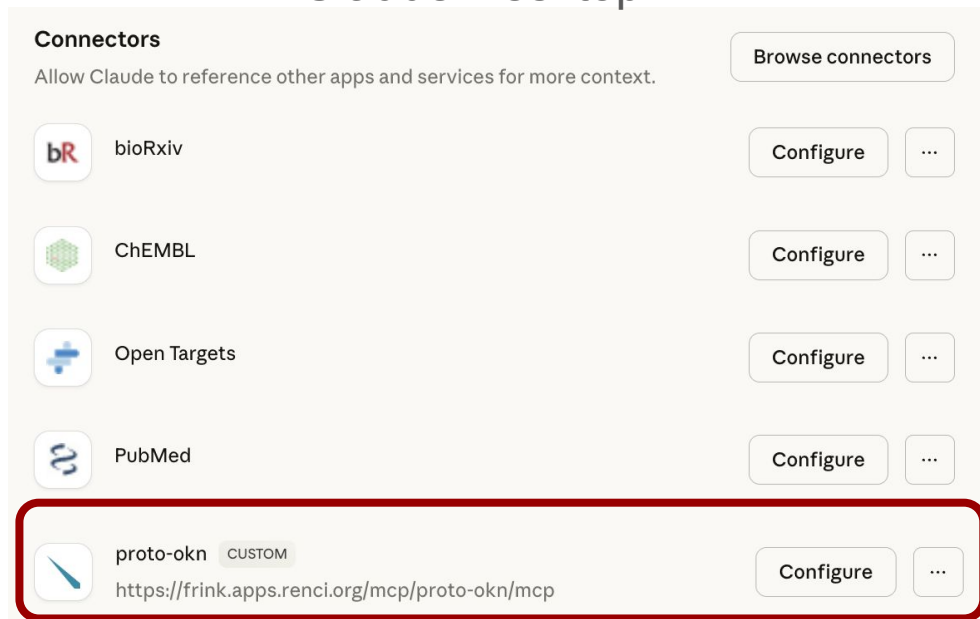
<https://modelcontextprotocol.io/>

MCP-Proto-OKN - Natural Language Queries



Setup mcp-proto-okn in AI Assistant

Claude Desktop

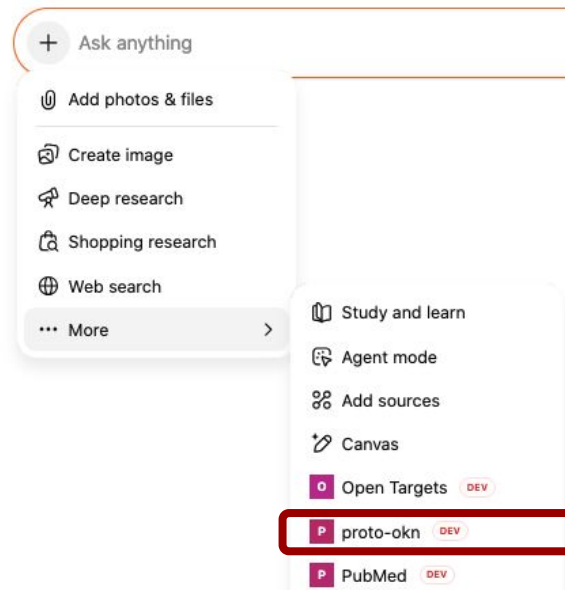


The screenshot shows the 'Connectors' section in Claude Desktop. It lists several connectors: bioRxiv, ChEMBL, Open Targets, PubMed, and proto-okn. The 'proto-okn' connector is highlighted with a red box. It is labeled as 'CUSTOM' and has the URL 'https://frink.apps.renci.org/mcp/proto-okn/mcp'.

Connectors
Allow Claude to reference other apps and services for more context.

proto-okn CUSTOM
https://frink.apps.renci.org/mcp/proto-okn/mcp

ChatGPT



The screenshot shows the ChatGPT interface with a list of actions. The 'proto-okn' connector is highlighted with a red box. It is labeled as 'DEV'.

proto-okn DEV

Configuration instructions:

<https://github.com/sbl-sdsc/mcp-proto-okn?tab=readme-ov-file#installation-and-configuration>

Demos

1. Proto-OKN KGs Descriptions

Generate a table of all Proto-OKN Knowledge Graphs with two columns: “KG Name” and “Description”

2. Single KG query

@spoke-genelab: Give a high-level overview of this knowledge graph, including its main entities, relationships, and purpose

3. Perfluorooctanoic acid (PFOA) safety profile - cross-KG

<https://github.com/sbl-sdsc/mcp-proto-okn/blob/main/docs/examples/pfoa-safety-profile.md>

4. Spaceflight disease risks - cross-KG + cross-MCP

<https://github.com/sbl-sdsc/mcp-proto-okn/blob/main/docs/examples/space-flight-disease-relationships.md>

5. Spaceflight microbiome investigation - cross-KG + cross-MCP

https://github.com/sbl-sdsc/mcp-proto-okn/blob/main/docs/examples/spaceflight_microbiome_cross-graph_investigation.md

6. Full list of examples

<https://github.com/sbl-sdsc/mcp-proto-okn?tab=readme-ov-file#example-queries>

Tips for Writing Better Prompts

- Be specific, provide context, and clearly state your goal
- Break complex tasks into steps
- Constrain the scope (e.g., specific KG) and anchor to known entities
- Ask for the schema to explain the KG's structure
- Request evidence, assumptions, or intermediate reasoning
- Redirect it when it goes off track and iterate
- Give clear feedback on plots or diagrams
- Ask it to improve your prompt
- Ensure web search is off
- Specify the desired output format (table, csv, plot, etc.)
- Use the prompt “Create chat transcript” to create a document with prompts, responses, and model version that you can save as a .md or .pdf file (see examples)