

Technical Review

MCP-Proto-OKN: Natural Language Queries of the Proto-OKN Knowledge Graphs

Peter Rose (pwrose@ucsd.edu)

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)

FRINK TEAM

Christopher Bizon (RENCI), Jim Balhoff (RENCI)

NSF Award #2333819

NSF Proto-OKN Program

**Proto-OKN**

[About](#) [Projects](#) [Presentations](#) [Events](#) [Proto-OKN PI Portal](#) [Contact](#)



Knowledge Graph Use Cases

**Biology &
Health**

**Environment**

**Justice**

**Technology &
Manufacturing**

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

Save Query

Run Query

RESULTS

50 results in 55.24s

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

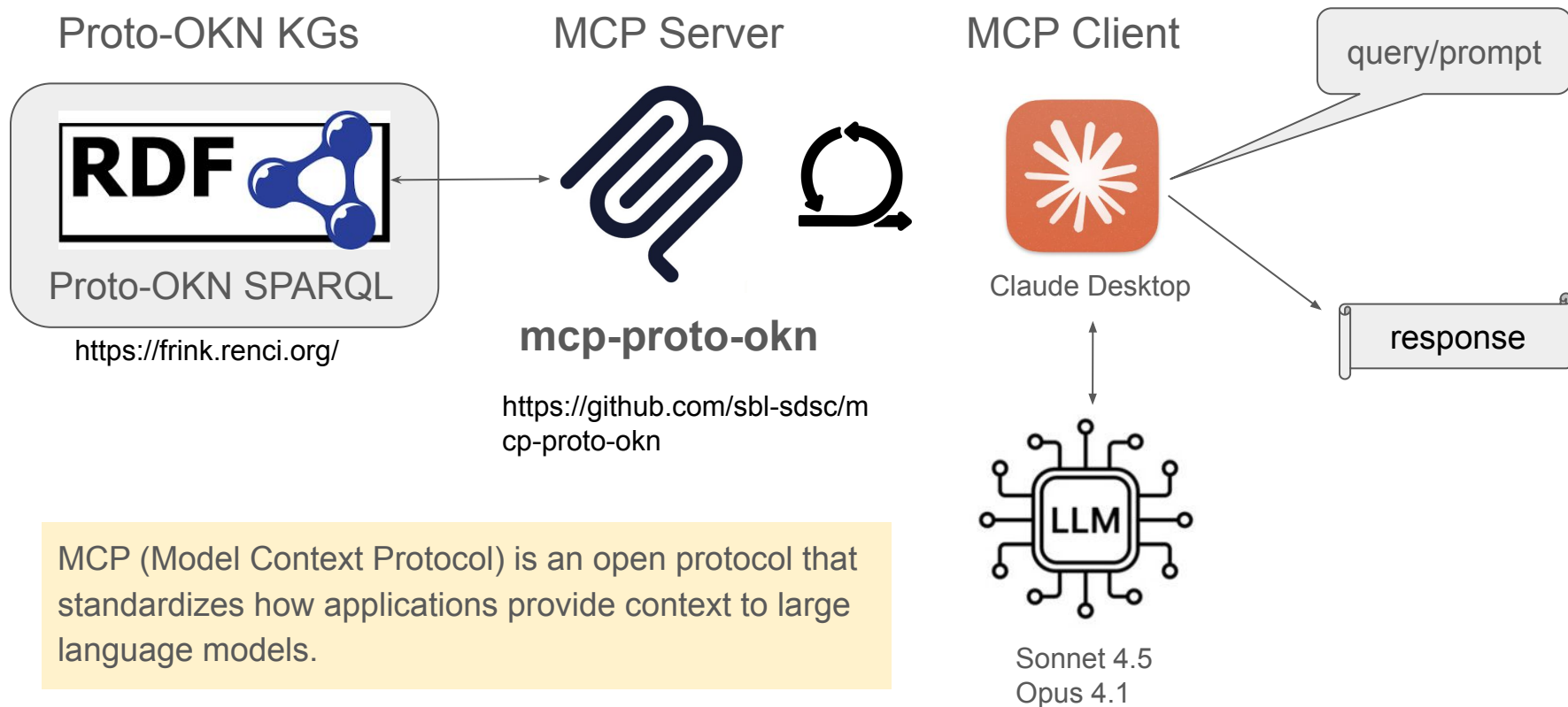
SPARQL
queries are
verbose and
complex

Few users
have
SPARQL
experience

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

Natural Language Queries

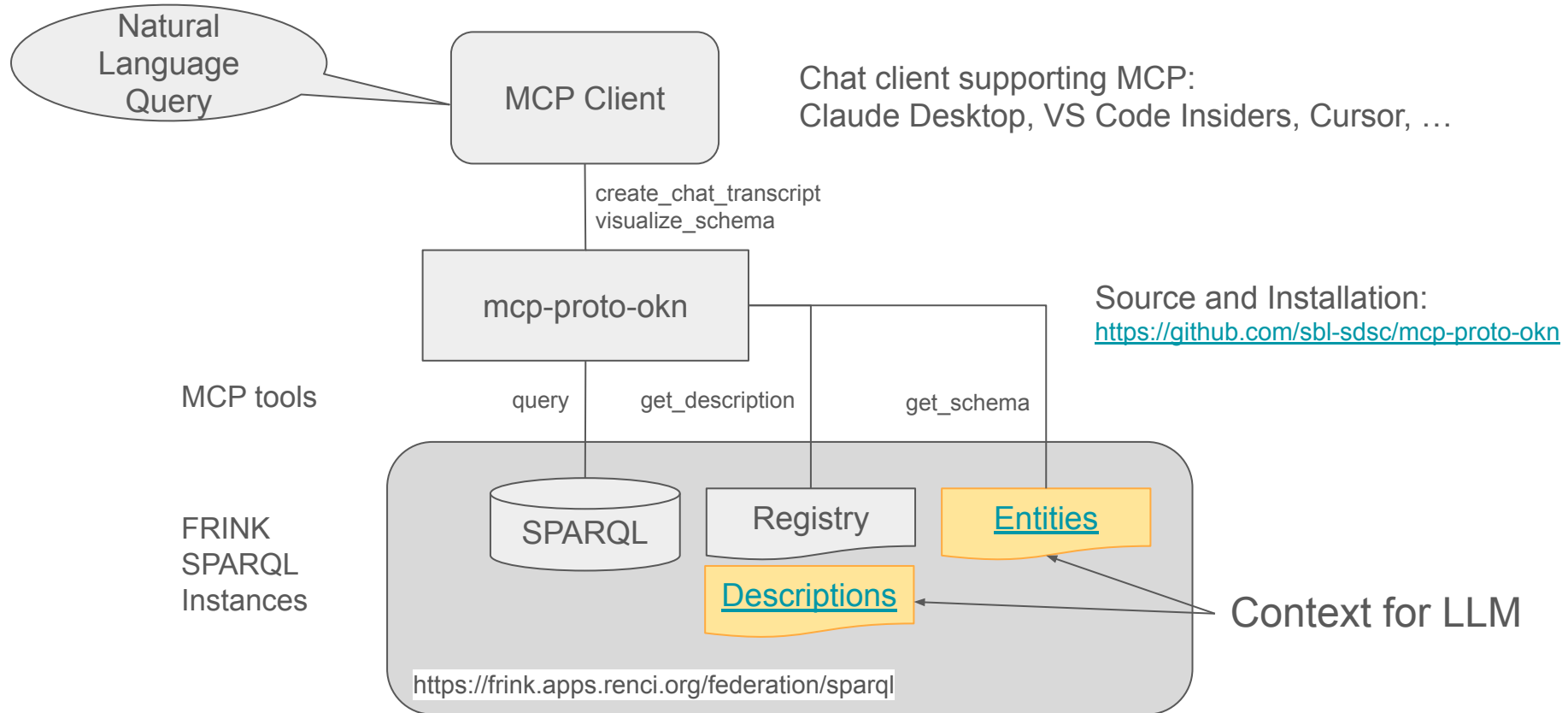
MCP-Proto-OKN Server Querying SPARQL



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

<https://modelcontextprotocol.io/>

MCP Architecture



Context Engineering

- KG Description

- ~150 word descriptions of the Proto-OKN KGs
 - Generated by **mcp-proto-okn** by analyzing the data in a KG and using publicly available information
- Descriptions guide the LLM to select the relevant KG(s) for a query.

- KG Entities

- We created a short label and description for each URI

URI	Label	Description
http://purl.obolibrary.org/obo/OBI_0000070	Assay	Unique identifier for an assay instance in the NASA Open Science Data Repository (OSDR)

- The LLM uses these descriptions to understand the content - Nodes, Edges, and their properties to transform natural language queries into SPARQL
- The entity descriptions were generated by **mcp-proto-okn** using the protocol described on the next slide

Descriptions and entities: <https://github.com/sbl-sdsc/mcp-proto-okn/tree/main/metadata>

Protocol to Generate the Entity Descriptions using mcp-proto-okn

For each <KG>, the following prompt was used to create the <KG>_entities file, where <KG> is the name of the KG in the FRINK registry:

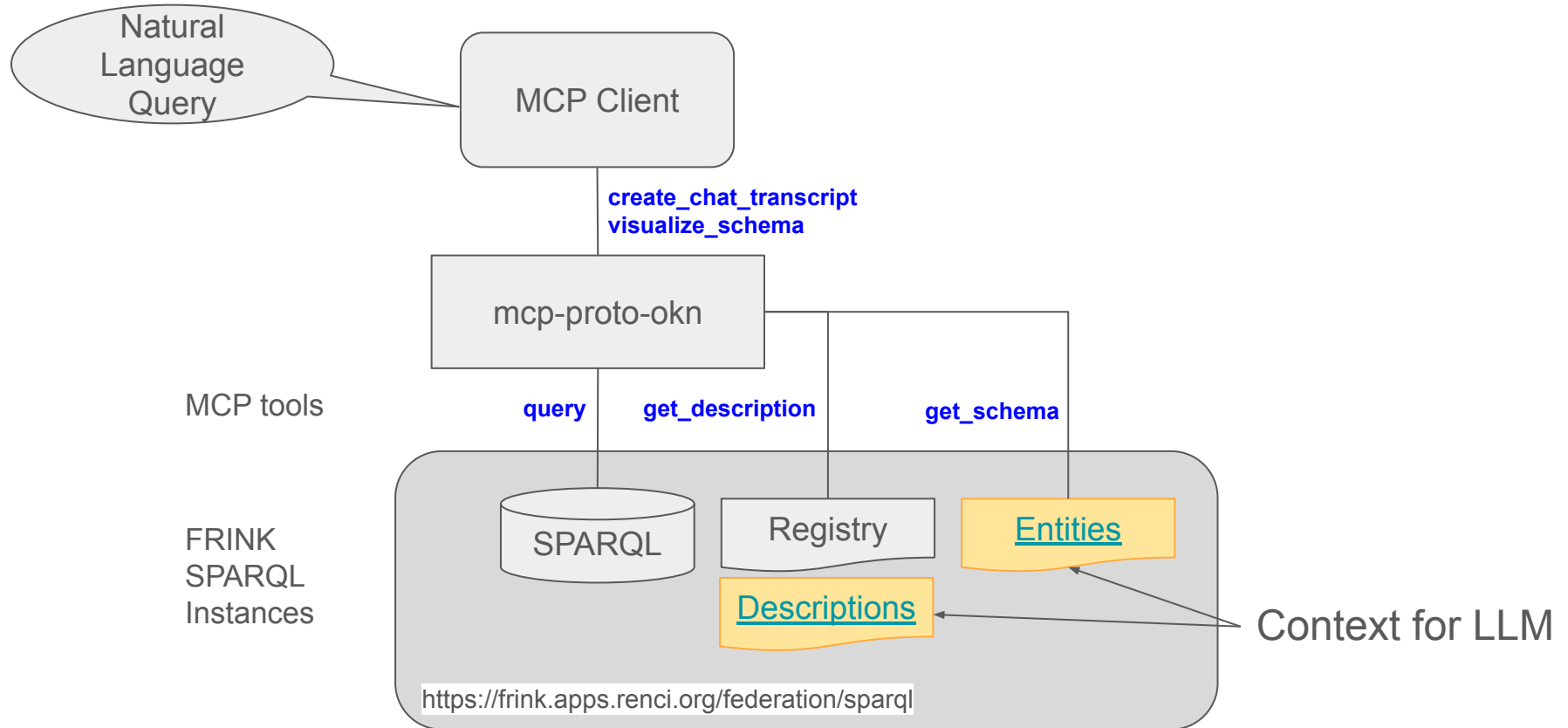
@<KG>: Create a csv file with the classes and predicates of this KG. The file should have 4 columns:

1. URI
2. Label - label of the URI
3. Description - description or comment about the URI
4. Type - either “Class” or “Predicate”

Use the get_schema tool to find the Classes and Predicates.

Search available ontology files, git repos, or other resources to retrieve this information. Add quotes to text with commas. Check the validity of the .csv file.

MCP-Proto-OKN Tools



mcp-proto-okn Tools

Tool permissions

✓ Always allow ▼

Choose when Claude is allowed to use these tools.

query



get_schema



get_description



get_query_template



Query tools

clean_mermaid_diagram



create_chat_transcript



visualize_schema



Visualization & reporting tools

Tool Descriptions: <https://github.com/sbl-sdsc/mcp-proto-okn/blob/main/docs/api.md>

Demos

Add MCP Server to Claude Desktop

Local MCP servers

Add and manage MCP servers that you're working on.

Edit Config

biobricks-aopwiki

biobricks-aopwiki running

biobricks-ice

Command

uvx

biobricks-mesh

Arguments

biobricks-pubchem...

mcp-proto-okn --endpoint https://frink.apps.renci.org/biobricks-aopwiki/sparql

biobricks-tox21

biobricks-toxcast

biohealth

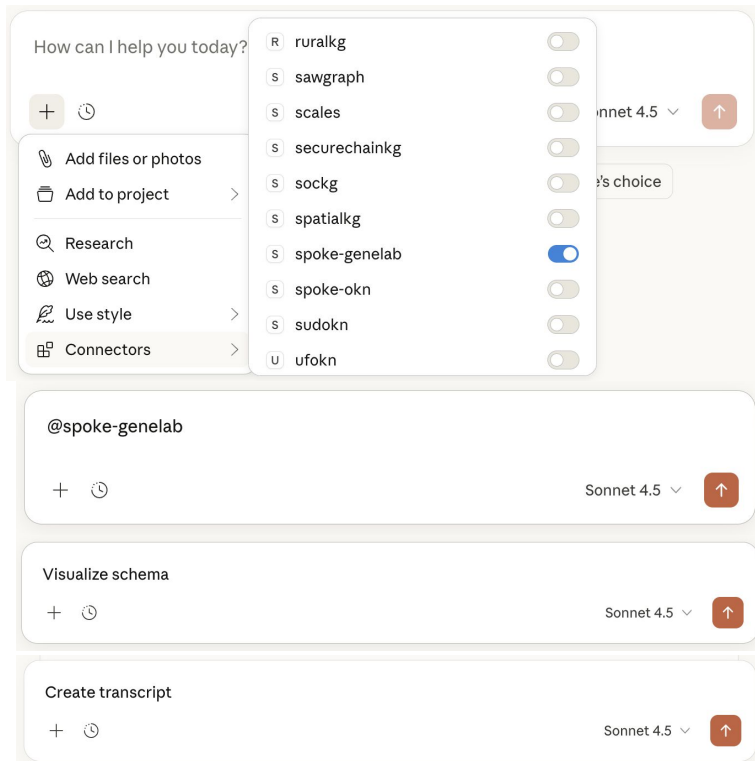
climatemodelskg

Instructions: 2 simple steps

<https://github.com/sbl-sdsc/mcp-proto-okn>

```
claude_desktop_config.json
{
  "mcpServers": {
    "biobricks-aopwiki": {
      "command": "uvx",
      "args": ["mcp-proto-okn", "--endpoint",
"https://frink.apps.renci.org/biobricks-aopwiki/sparql"]
    },
    "biobricks-ice": {
      "command": "uvx",
      "args": ["mcp-proto-okn", "--endpoint",
"https://frink.apps.renci.org/biobricks-ice/sparql"]
    },
    "biobricks-mesh": {
      "command": "uvx",
      "args": ["mcp-proto-okn", "--endpoint",
"https://frink.apps.renci.org/biobricks-mesh/sparql"]
    },
  }
}
```

Demo 1 - spoke-genelab Overview and Schema



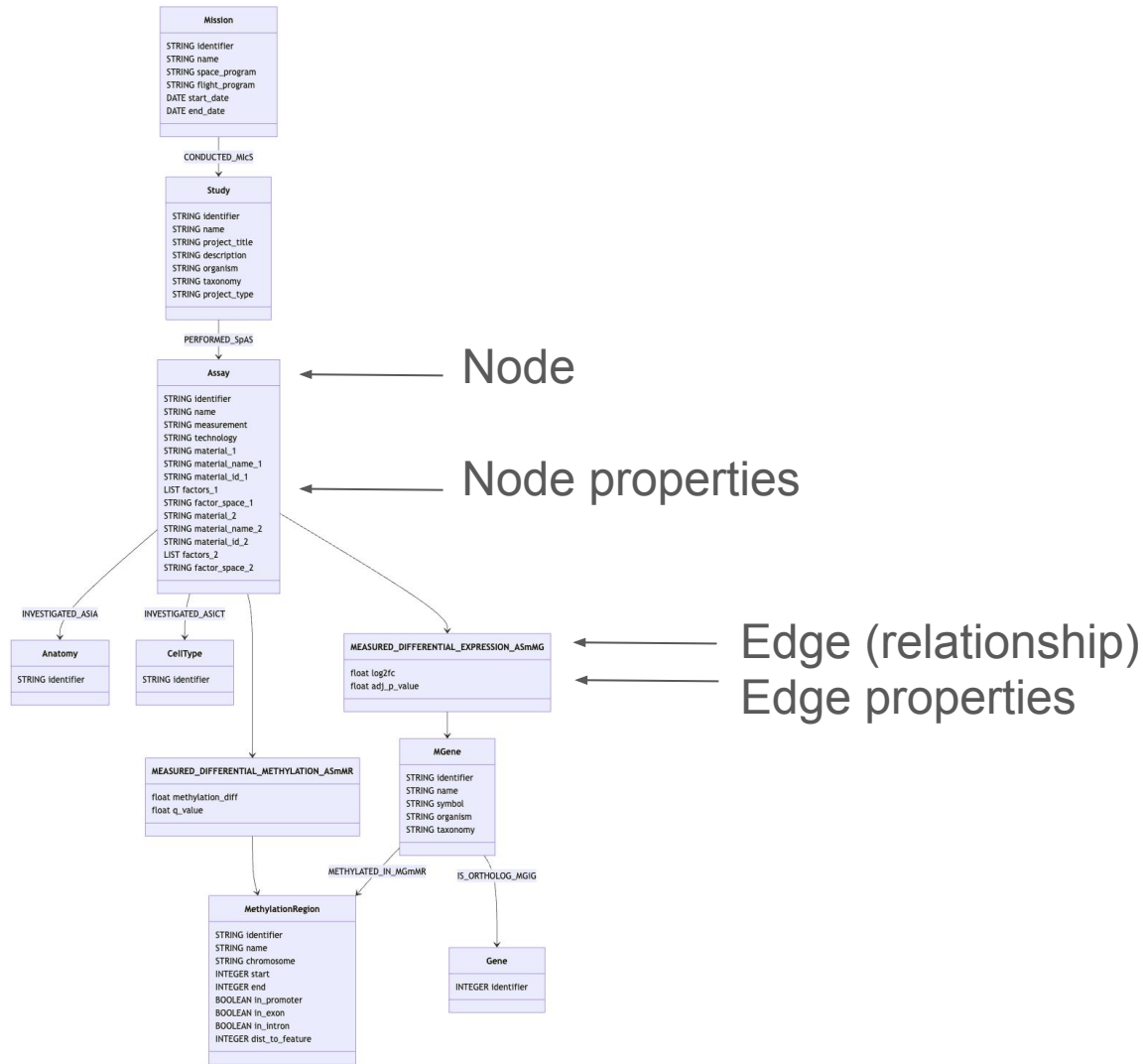
Select KG(s) to query

1. Prompt: describe spoke-genelab KG
2. Prompt: generate the KG schema diagram
3. Prompt: create a transcript (.md, .pdf)

Transcript:

https://github.com/sbl-sdsc/mcp-protocol-okn/blob/main/docs/examples/spoke-genelab_overview.md

LLM-generated spoke-genelab KG Schema



Demo 1

What type of data are in the Proto-OKN KGs?

Knowledge Graph Overviews & Class Diagrams

Each link below points to a chat transcript that demonstrates how to generate a knowledge-graph overview and class diagram for a given Proto-OKN Theme 1 KG. The examples are grouped by domain area.

 Biology & Health	 Environment	 Justice	 Technology & Manufacturing	Other
biobricks-aopwiki	sawgraph	ruralkg	securechainkg	nasa-gesdisc-kg
biobricks-ice	fiokg	scales	sudokn	
biobricks-mesh	geoconnex	nikg		
biobricks-pubchem-annotations	spatialkg	dreamkg		
biobricks-tox21	hydrologykg			
biobricks-toxcast	ufokn			
spoke-genelab	wildlifekn			
spoke-okn	climatemodelskg			
	sockg			

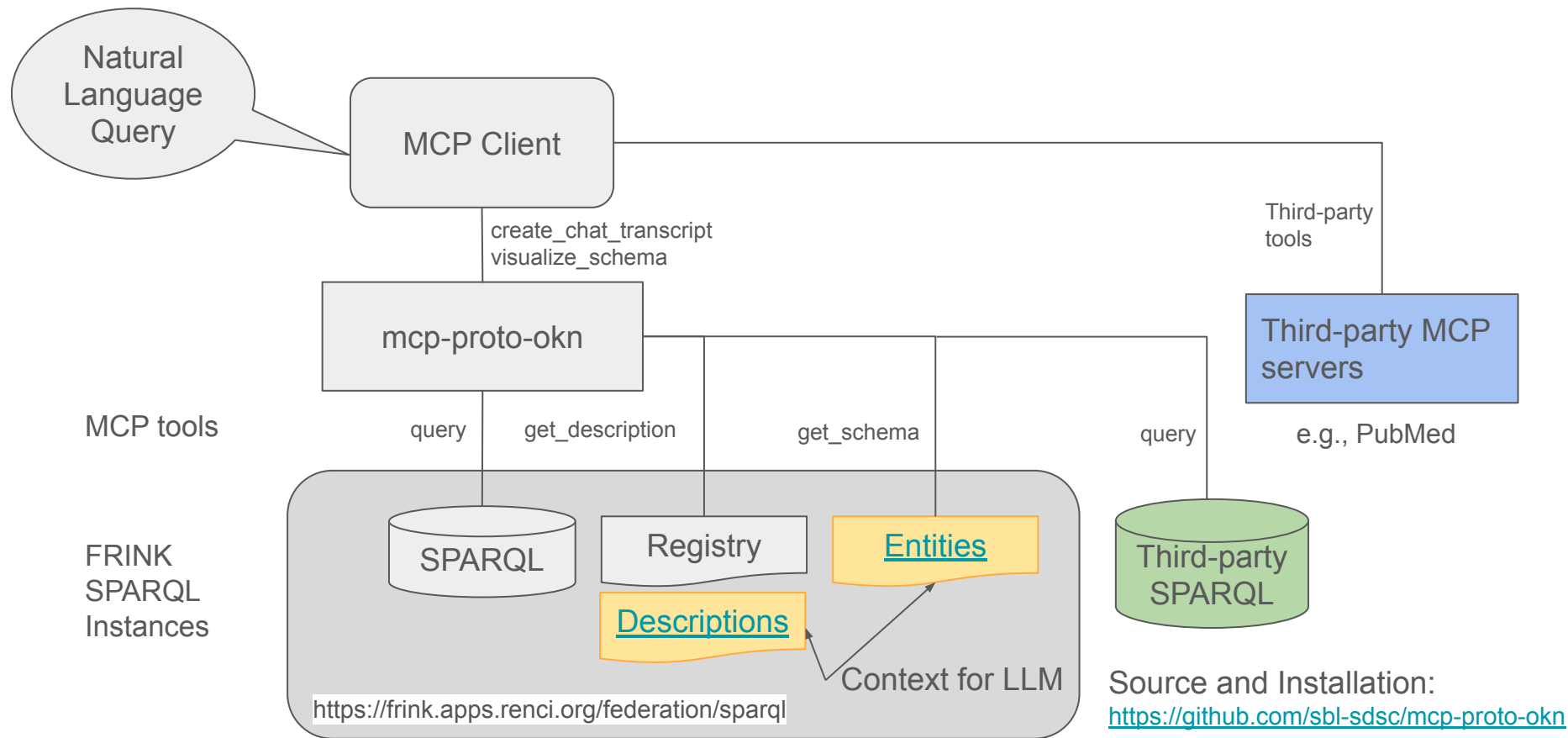
Transcripts of Proto-OKN KGs with overviews and schemas are available.

<https://github.com/sbl-sdsc/mcp-proto-okn/tree/main?tab=readme-ov-file#knowledge-graph-overviews--class-diagrams>

Demo 2

Queries across multiple Proto-OKN KGs

Extending the Proto-OKN Ecosystem: Third-party Integration



Demo 2 - Differential Gene Expression Analysis

Integrate GeneLab, SPOKE-OKN, and PubMed knowledge to characterize spaceflight-associated gene expression changes and their disease relevance for GeneLab study OSD-161.

Transcript:

<https://github.com/sbl-sdsc/mcp-proto-okn/blob/main/docs/examples/osd-161-sonnet-4.5.md>

Demo 3: Use Cases for several Proto-OKN KGs

Use Cases

1. [Spaceflight Missions \(spoke-genelab\)](#)
2. [Spaceflight Gene Expression Analysis \(spoke-genelab, spoke-okn\)](#)
3. [Spaceflight Gene Expression with Literature Analysis \(spoke-genelab, spoke-okn, PubMed\)](#)
4. [Disease Prevalence in the US \(spoke-okn\)](#)
5. [Disease Prevalence - Socio-Economic Factors Correlation \(spoke-okn\)](#)
6. [Contamination at Superfund Sites \(spoke-okn\)](#)
7. [PFOA in Drinking Water \(spoke-okn\)](#)
8. [Data about PFOA \(spoke-okn, biobricks-toxcast\)](#)
9. [Biological Targets for PFOA \(biobricks-toxcast\)](#)
10. [Criminal Justice Patterns \(scales\)](#)
11. [Drug Possession Charges \(scales\)](#)
12. [Environmental Justice \(sawgraph, scales, spatialkg, spoke-okn\)](#)
13. [Rural Health Access \(ruralkg, dreamkg, spoke-okn\)](#)
14. [Michigan Flooding Event \(ufokn\)](#)
15. [Flooding and Socio-Economic Factors \(ufokn, spatialkg, spoke-okn\)](#)
16. [Philadelphia Area Incidents \(nikg\)](#)

<https://github.com/sbl-sdsc/mcp-protokn/tree/main?tab=readme-ov-file#use-cases>

Demo 4: Proto-OKN Integration Opportunities

Cross-KG Geolocation Analysis Summary

1. Geolocation Data Inventory Highlights

Richest Geolocation Resources:

- **SAWGraph Spatial KG:** 7.4M S2 cells (Level 13) plus ~42K administrative regions forming the backbone spatial infrastructure
- **SAWGraph FRS KG:** 5.7M EPA-regulated facilities with 4.4M S2 cell connections
- **Geoconnex:** 1.15M point geometries plus 34K line features for hydrologic locations
- **SPOKE-OKN:** 246K WKT geometries covering ZIP codes, counties, and international locations

Key Geolocation Types Found:

Type	Primary KGs	Count Range
S2 Cells (Level 13)	Spatial, FRS, Hydrology, SOC-KG	7.4M
Point Geometries (WKT)	Geoconnex, NIKG, SOC-KG	1.2M+
Administrative Regions	Spatial KG, Rural KG, SPOKE-OKN	45K+
Named Locations	Climate Models, SUDOKN	55K+
FIPS/ZIP Codes	Multiple KGs	40K+

Details: <https://github.com/sbl-sdsc/mcp-proto-okn/tree/main?tab=readme-ov-file#proto-okn-integration-opportunities>

Cross-Platform LLM Benchmarks (as of Jan. 5, 2026)

Claude Desktop

Opus 4.5

Most capable for complex work

Sonnet 4.5

Best for everyday tasks

Haiku 4.5

Fastest for quick answers

More models



VS Code Insiders/GitHub Copilot

Claude Haiku 4.5	0.33x
Claude Opus 4.5	3x
Claude Sonnet 4	1x
✓ Claude Sonnet 4.5	1x
Gemini 2.5 Pro	1x
Gemini 3 Flash (Preview)	0.33x
Gemini 3 Pro (Preview)	1x
GPT-5	1x
GPT-5-Codex (Preview)	1x
GPT-5.1	1x
GPT-5.1-Codex	1x
GPT-5.1-Codex-Max	1x
GPT-5.1-Codex-Mini (Previ...	0.33x
GPT-5.2	1x

Demo 5: Cross-Platform LLM Benchmarks

This section compares the results of two queries using Claude Desktop and VS Code Insiders with commons LLMs.

Query	Claude Sonnet 4.5	Claude Sonnet 4.5	Gemini 3 Pro	Groq Code Fast 1	GPT-5.2
Spaceflight Missions	Claude Desktop	VS Code	VS Code	VS Code	VS Code
Gene Expression Analysis	Claude Desktop	VS Code	VS Code	VS Code	VS Code

Evaluation: Claude Sonnet 4.5 > Gemini 3 ~ GPT-5.2 > Groq Code Fast 1

Details: <https://github.com/sbl-sdsc/mcp-protocol/tree/main?tab=readme-ov-file#cross-platform-llm-benchmarks>

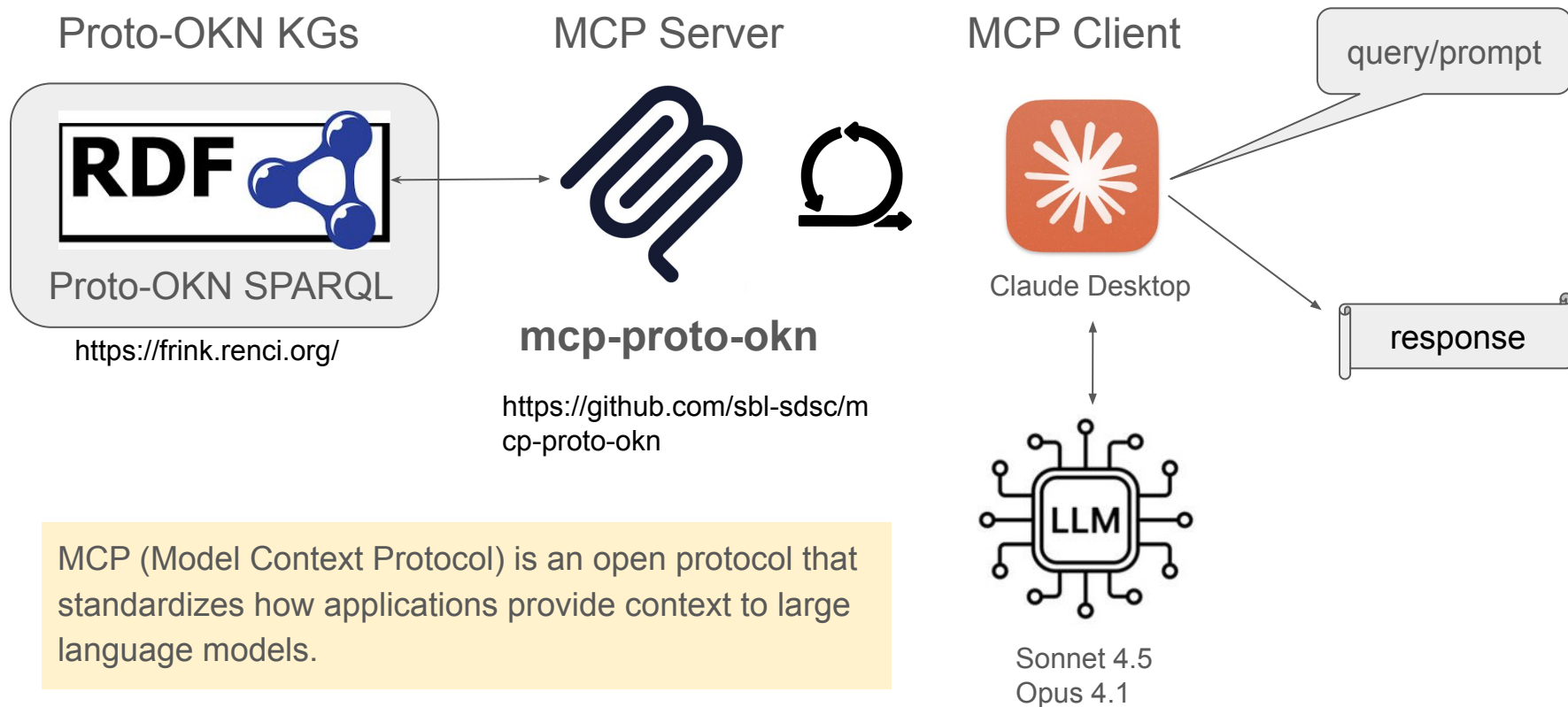
Leveraging the Model Context Protocol (MCP) for Integration and Query of Proto-OKN Knowledge Graphs

- Unified Natural Language Interface: Enables users to query multiple Proto-OKN KGs using natural language via clients like Claude, or VS Code.
- Modular Server Integration: Connects to multiple mcp-proto-okn modules, each handling SPARQL queries to specific FRINK endpoints, allowing scalable and flexible architecture.
- Federated Knowledge Access: Supports querying across distributed knowledge graphs, identifying relationships and integration points across domains.
- Automated Knowledge Discovery: Facilitates extraction and summarization of content across KGs, and semantic alignment without explicit instructions.
- Interoperability Hub: Acts as a middleware to integrate third-party resources with Proto-OKN KGs, enabling broader data harmonization and reuse for scientific and technical exploration.

Plans

- Develop Use Cases in Collaboration with Proto-OKN teams
- Establish Benchmarks for the mcp-proto-okn server
 - Develop benchmark datasets for evaluating natural language queries
 - Optimize MCP server using performance results from benchmarking
- Advance Knowledge Graph Interoperability
 - Demonstrate federated queries across Proto-OKN Knowledge Graphs
 - Incorporate multi-agent queries that leverage third-party KGs and MCP servers

MCP-Proto-OKN Server Querying SPARQL



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

<https://modelcontextprotocol.io/>