



JOHNS HOPKINS
UNIVERSITY

Bloomberg Center for
Government Excellence

5 October 2026

Building a Regulatory Knowledge MCP

(How we learned to stop worrying and love the AI)

A decorative graphic at the bottom of the slide consisting of several overlapping, wavy lines in a light blue color, creating a sense of motion and depth.

25 minutes to talk about...

What are we trying to do?

Assembling the sources

Building the data store

Tools of the MCP

Discussion

Optimizations and other findings (time permitting)

Problem: City Municipal Regulations Accessibility

City Code and Municipal Regulations Cover?

- Business Licensing
- Real Estate Rental Rules
- Utilities Regulations
- Agricultural restrictions– can I have a chicken coop in my backyard? If I can, *can I eat them??*
- Education and Childcare Requirements
- Basically every aspect of life in a city is touched on by the municipal code



Problem: City Municipal Regulations Accessibility

Resident Query

“In the city, are daycare centers required to follow Safe Sleep practices for infants?”

Web Search Result

Ambiguous answers with a lack of traceability back to official sources.

This parents goal is aligned with business owners, developers, city staff, and policy researchers

- Answers that preserve provenance of information
- Clarity in the face of complex primary sources



Photo by Ajay Suresh
<https://flic.kr/p/2m6JZ5g> CC License

Problem: City Municipal Regulations Accessibility

Resident Query

“In the city, are daycare centers required to follow Safe Sleep practices for infants?”

Direct From City Municipal Records Website

Search Term	Results
<i>daycare sleep rules</i>	0
<i>nap time</i>	0
<i>infant sleep</i>	20
<i>cribs</i>	52

There are two documents in these last sets. 1 is a nice half a page explanation of the rules on how infants should sleep. The other is a 62,000 word monstrosity.



Photo by Ajay Suresh
<https://flic.kr/p/2m6JZ5g> CC License

Proposed Solution

GRAPH RAG Powered MCP Service

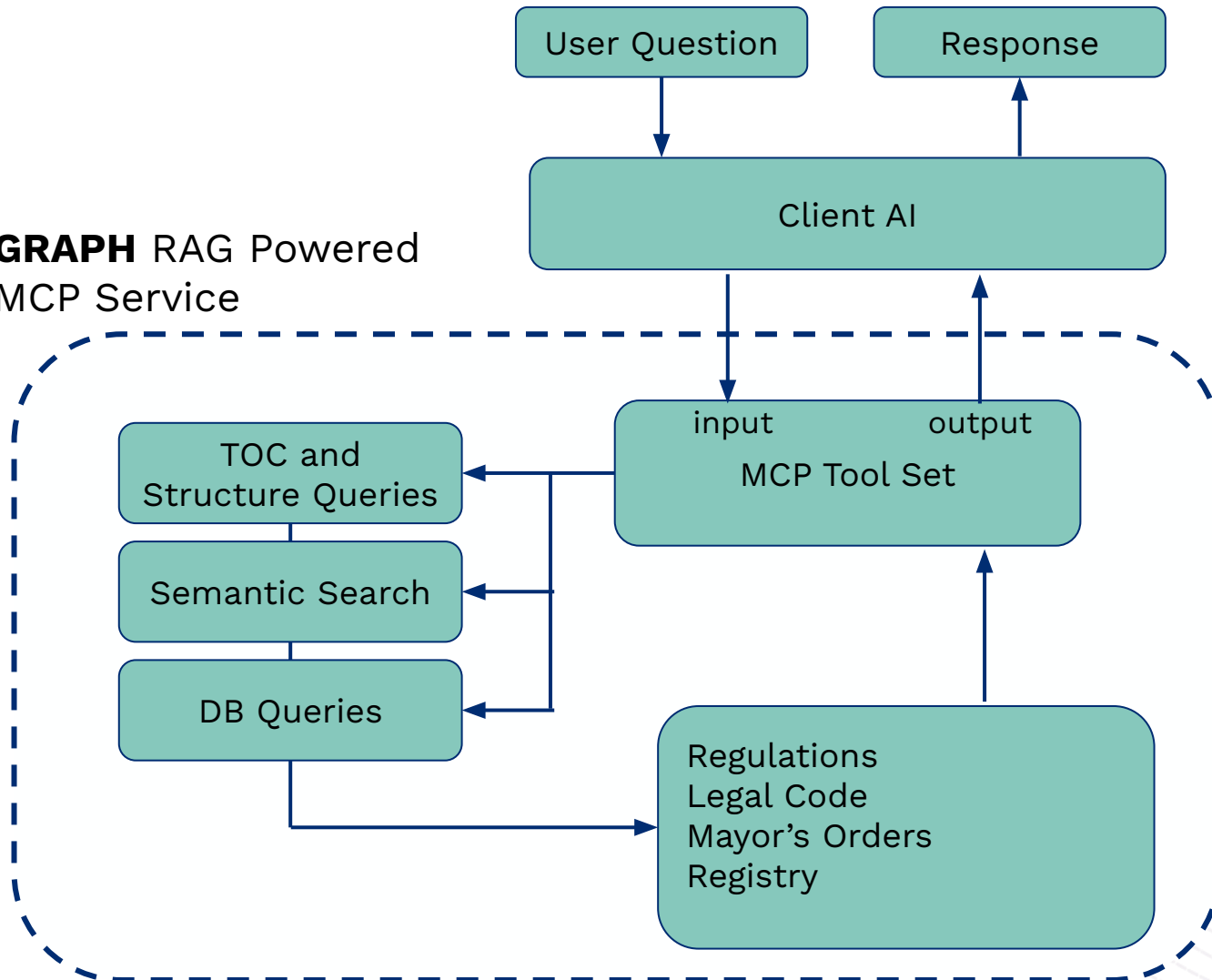


Photo by Ajay Suresh
<https://flic.kr/p/2m6JZ5g> CC License

Data Landscape

The data is only a minor nightmare

Municipal Regs

24,000+ sections and each one is a Word document.

Code of City of <Redacted>

Parse a lot of xml. At least it's structured.

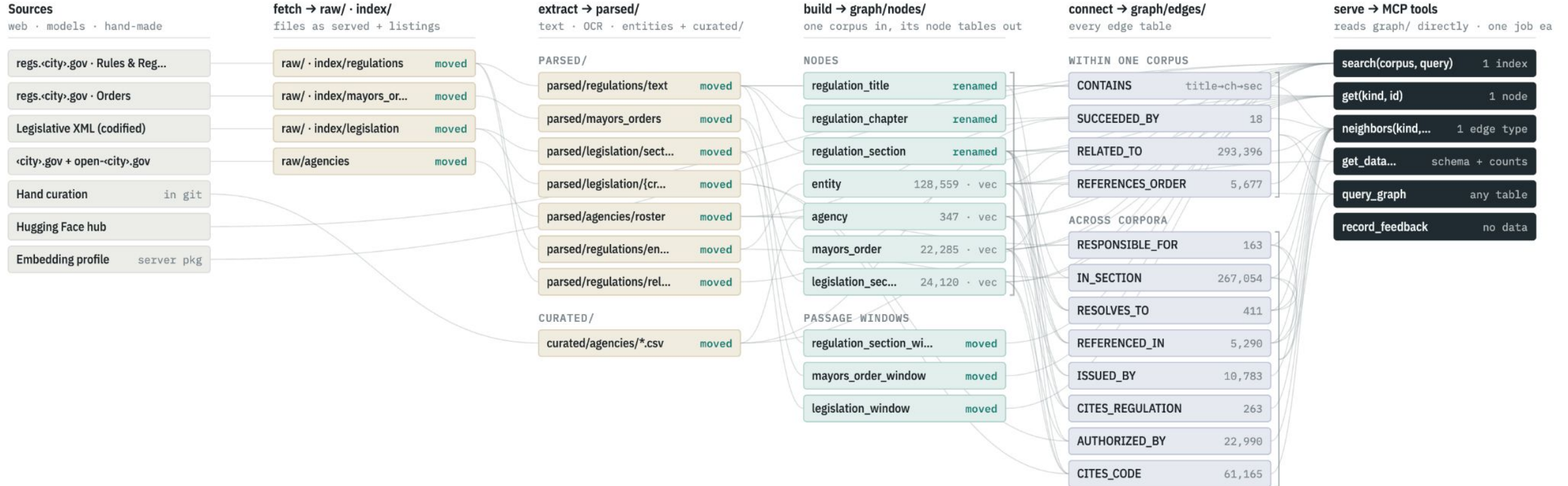
Mayor's Orders

A pile of PDFs, with terrible text layer quality

What we can do with our MCP server depends on how we process, analyze, and transform this data.

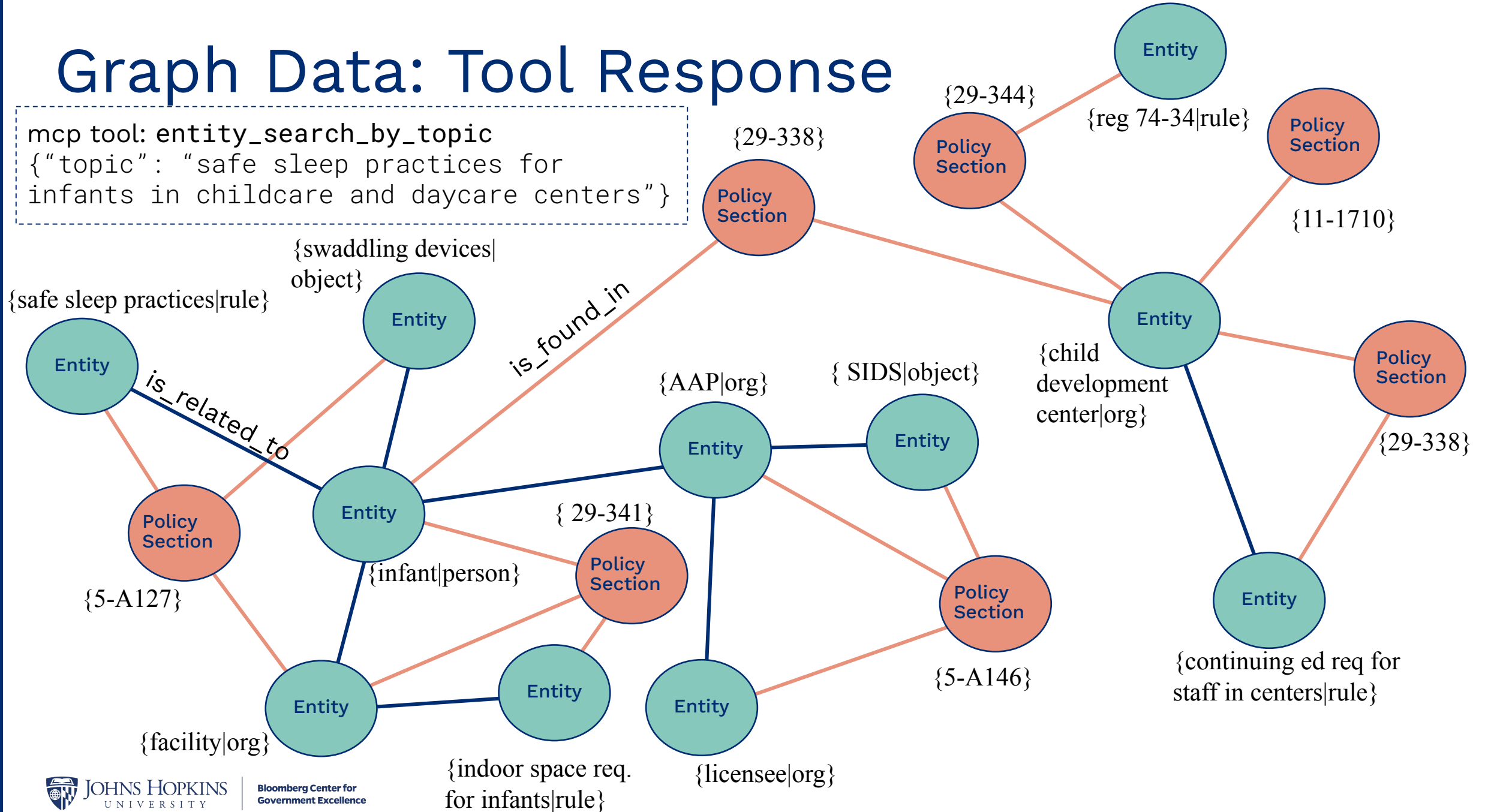


Pipeline: Data Extraction



Graph Data: Tool Response

mcp tool: entity_search_by_topic
{“topic”: “safe sleep practices for infants in childcare and daycare centers”}



Producing a Response

Initial Tool Call

Entity Query:

- 37 entities
- 45 relationships

Section Vector Search:

- 30 sections

Short descriptions and text excerpts allow client LLM to reason through relevant connections and identify the sections necessary to identify the answer

Entity Example

```
{'entity': 'FACILITY',  
  'entity_type': 'ORGANIZATION',  
  'direction': 'out',  
  'description': 'Safe sleep practices are followed  
by the center (facility).',  
  'via_section_id': '5-A127'}
```

Relationship Example

```
{'entity_id': 'SAFE SLEEP PRACTICES|RULE',  
  'title': 'SAFE SLEEP PRACTICES',  
  'type': 'RULE',  
  'description': 'Guidelines followed by the facility to  
ensure a safe sleep environment for infants, including  
specific instructions on infant placement and items  
allowed in sleeping areas.',  
  'via_section_id': '5-A127',}
```

Section Example

```
{'section_id': '5-A127',  
  'regs_title': '5',  
  'regs_chapter': '5-A1',  
  'regs_section': 'A127',  
  'body_excerpt': '... procedures;\n\n) A description of the  
safe sleep practices followed by the Center that includes  
the following information:\n\n1) When setting an infant  
down to sleep, the infant will be placed ...'}
```

Producing a Response

Assembling an Answer

Document Direct Lookup (only 4!):

- **5-A146:** *Health, Safety, and Welfare: Safe Sleeping and Resting Practices*
- **29-376:** *Requirements for Infant Sleep and Play Positions (Precautions against SIDS)*
- **5-A127:** *Parent Communication*
- **29-365:** *Cots, Cribs, and Infant Play*

The knowledge graph condenses explanations of the document relationships and contents allowing the client LLM to identify the most relevant documents and anticipate what to look for in each in order to construct a response.

Answer (shortened)

Yes — and it's one of the most explicit requirements in the MR. Any licensed child development facility in the City that cares for one or more infants must follow safe-sleep practices.

[excerpt describing them with citations]

*§ 5-A146 (Health, Safety and Welfare: Safe Sleeping and Resting Practices) binds every licensee caring for infants, and makes **AAP** guidance the floor:*

[excerpt describing relationship between regulations and AAP recommendations]

Coverage: **which facilities**

The duty attaches to every facility type that cares for infants — the 5-A chapter's own structure shows it reaches child development centers (§ 5-A162), homes (§ 5-A167), and expanded homes (§ 5-A169).

The documentation obligation: § 5-A127

*Safe sleep isn't just operational — it must be written down and shared with parents. § 5-A127 (**Parent Communication**) requires every licensee to "establish and implement written policies and procedures"*

Big questions we have

- Trust: what makes our service - the official one for the municipality, trustworthy? Especially with the proliferation of registries?
 - Ex: there is an official US Census MCP ... and an unofficial one (which might be better)
 - Which platform registries? How do we prove officialness?
- There are 88,000 government entities in the U.S. How do we scale effectively?
- Long term stability is important to governments
 - How do we make sure we don't leave older LLMs behind?
 - What kinds of resources are needed for services to remain valuable?

Closing Thoughts

We are planning on an eventual public release of at least parts of this tool.

City partner will be piloting this testing this with staff soon.

MCP servers allow for creative and flexible data access. This is just one possibility.

Graph data is awesome.

Optimizations and other findings

MCP Design & LLM Behavior

- It's VERY HARD to separate MCP from harness from LLM, and each multiplies the unpredictability / uncertainty / imprecision
- A purpose-built agent with the service info in the system prompt is way more effective than coaching the LLM through service overview & tool descriptions
- Formatting content/text into markdown rather than a stringified JSON copy of structuredContent makes a world of difference, especially to older agent clients
- We can build a feedback mechanism right into the MCP!

Optimizations and other findings

Retrieval & Storage

- Reducing token consumption: 2-stage vector search: first pass finds best matching document(s), second finds best matching text snippet(s) within those document(s)
- Using 512t-input embedding + chunking is more effective than 8192t-input; further improved by combining doc & chunk distances
- Matryoshka Representation Learning (MRL) embeddings & truncated/normalized vectors reduced storage by ~60% (768d -> 256d)
- Adding raw graph querying enabled LLMs to do quantitative work as well as more sophisticated content analyses
 - Ex: What kinds of regulations typically cite repealed / replaced laws, and which agencies are most affected?

Optimizations and other findings

Architecture & Stack

- Python + FastMCP 3.x
- LadybugDB: a graph store w/ vector search extension support in the style of SQLite & ChromaDB
- Using graph network traversal is faster than "SELECT * FROM Nodes WHERE ..."

Content Ingestion

- Industrial & off-the-shelf OCR is awful, specialized VLMs are the way to go!
- Tables in documents are still a problem; retrieval at row x column y has ~50-60% accuracy for the best LLMs

Thank you!

Maeve Mulholland

maeve.mulholland@jhu.edu

Andrew Nicklin

anicklin@jhu.edu



JOHNS HOPKINS
UNIVERSITY

**Bloomberg Center for
Government Excellence**

