

2024-03-05 DMRWG Meeting Notes

Meeting Date

23 Jan 2023 The DMRWG meets bi-weekly on Tuesdays at 12:00-13:00 PT / 16:00-17:00 UTC. Check the [ToIP Calendar](#) for meeting dates.

Zoom Recording & supporting material

- [Recording](#)

Attendees

- [Neil Thomson](#)
- [Steven Milstein](#)
- [Carly Huitema](#)
- [Burak Serdar](#)

Main Goal of this Meeting

Discussion on integrating data from relational and other non-vector databases into LLM AI models, via DAG and other techniques

Agenda Items and Notes (including all relevant links)

Ti me	Agend a Item	L e ad	Notes

5 m in	<ul style="list-style-type: none"> • S t a r t r e c o r d i n g • W e l c o m e & a n t i t r u s t n o t i c e • In t r o d u c t i o n o f n e w m e m b e r s • A g e n d a r e v i e w 	C h a i r s	<ul style="list-style-type: none"> • Antitrust Policy Notice: <i>Attendees are reminded to adhere to the meeting agenda and not participate in activities prohibited under antitrust and competition laws. Only members of ToIP who have signed the necessary agreements are permitted to participate in this activity beyond an observer role.</i> • New Members:
5 5 m ins	Discus sion	A ll	<p>Neil Thomson began the conversation by discussing the importance of understanding certain topics, expressing uncertainty about the extent to which Barack was prepared to discuss today. He mentioned that Anita Rao and Savida had talked about gathering some information on AI, suggesting a deeper dive into the subject might be needed, possibly in a separate session or during the AI discussion. Neil noted Wen Jing's recent travels in Europe and his involvement in writing on the trust spanning protocol, speculating on Wen Jing's current engagement with AI-related work. Neil then invited Stephen to share his opinions.</p> <p>Steven Milstein responded, indicating he had posted in the chat about the complexity of SQL and structured data, based on an article that caught his attention. He shared his experience with metadata and systems generating their own metadata upon content ingestion, allowing for SQL queries on such data. Steven elaborated on the usefulness of databases and discussed the nuances of querying language models, highlighting the differences in search scope and the unpredictable nature of AI-generated responses.</p> <p>Neil Thomson then shifted the conversation to the integration of databases, assuming these databases have a schema that defines their structure. He expressed interest in how databases could co-present or be queried for information. Steven and Neil discussed the potential of using APIs to access databases and the simplicity of using chat GPT for querying without writing any API code.</p> <p>The conversation took a turn towards data analytics, with Neil sharing insights from a BI analytics background. He talked about the corporate world's reliance on repeatable processes and the challenge of incorporating AI-generated insights into such a framework. Carly Huitema joined the discussion, emphasizing the human-like unpredictability of AI responses and the iterative nature of refining queries to get desired outcomes.</p> <p>Neil further explored the concept of a traveler profile, emphasizing the need for it to be human-understandable and verifiable. The discussion broadened to include selective disclosure of personal information, the travel industry's role in managing personal data, and the challenges of maintaining a continuous, selective disclosure framework.</p> <p>Burak Sendar shared his experience with Neo4j and the different approaches of using structured data to create knowledge graphs, contrasting it with how LLMs generate semantic models from language. He highlighted the conceptual gap between traditional databases and LLMs, and the potential but challenging integration of Neo4j data with LLMs to generate more insightful answers.</p> <p>Steven and Burak discussed the possibility of embedding Neo4j into a vector store for LLMs, speculating on the outcomes and acknowledging the conceptual differences between the two technologies.</p> <p>The transcript concluded with further discussion on the potential applications of Neo4j and LLMs in generating knowledge graphs and the ongoing challenges of data privacy and consent in the use of personal data for AI and analytics.</p>

Supporting Material:

[Human Colossus Foundation](#)

- [Overlay Capture Architecture](#) (home page for supporting material)

[Cloud Privacy Labs](#)

- [Presentation - interoperability with Layered Schemas](#)
- [JSON-LD vs. Layered JSON Schemas](#)
- [Demonstration of Layered Schema Architecture](#)
- [On Semantic Interoperability and Self-Describing Data](#)
- [Layered Schema Architecture](#) open source site

Screenshots/Diagrams (numbered for reference in notes above)

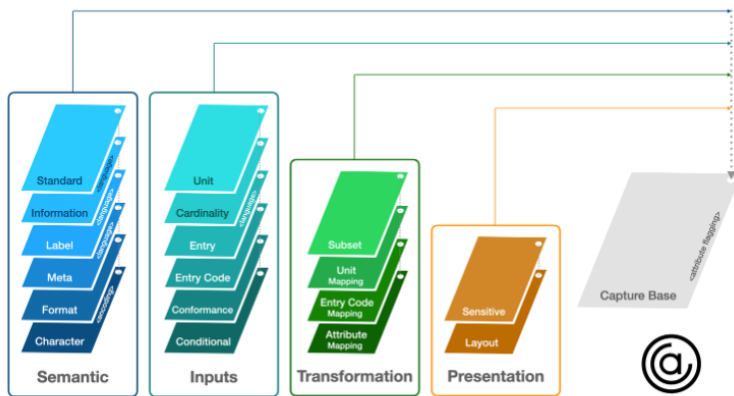


Figure 1. *Overlays Capture Architecture (OCA)* offers a data harmonisation solution within and across data ecosystems.

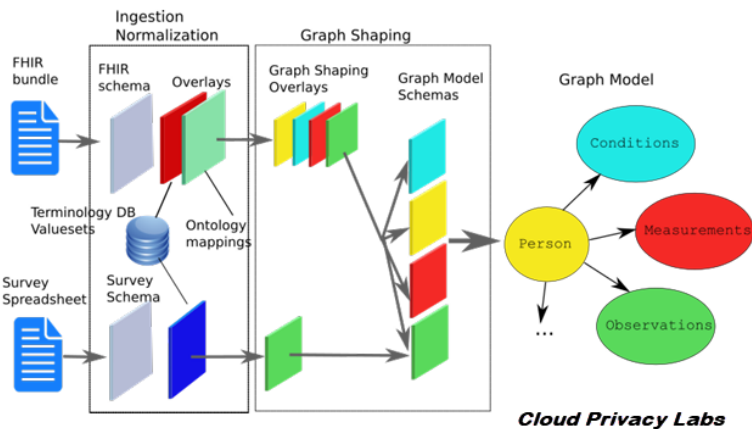


Figure 1: Pipeline to ingest FHIR and SDOH survey data into a common graph model

Decisions

- Sample Decision Item

Action Items

- ☐ Sample Action Item

