Leveraging a Semantic Layer for Research Curation and Conversational Experiences

The Challenge

A global philanthropic organization focused on health programs struggled to fully leverage knowledge from semi-structured and unstructured documents. Specifically, within a health-related funding program, researchers lacked access to key qualitative data from end-user surveys and transcripts. Consequently, they were unable to fully incorporate end-user experiences into their decision-making processes. This blind spot cost the organization significant time and resources, as researchers manually generated artifacts that lacked critical context.

At an enterprise level, multiple teams faced difficulties ingesting the complex, unstructured knowledge assets required for research or investment strategy. Without a standardized process, the organization wasted resources developing disconnected projects across programs. In addition, the enterprise sought guidance on how to utilize a newly acquired suite of advanced AI and semantic tooling, as well as how to govern and apply best practices to their knowledge and data management efforts.

The Solution

EK partnered with the organization to develop end-user capabilities for the health program and establish a repeatable pattern to scale these capabilities across enterprise programs and applications. To demonstrate the pattern's effectiveness, EK first piloted the capabilities within a prioritized program.

Discovery & Use Case Backlog:
EK initiated a series of modeling discovery workshops with the health program department to identify a starting point, key data challenges, and a repeatable use case. The team developed personas to map the data and process challenges facing researchers, translating these requirements into a backlog of potential AI use cases to pilot. After narrowing the focus with the health program team to qualitative product research insights, EK formalized a set of business outcomes and requirements that would guide the overall solution.

Domain Taxonomies & Ontology:
A core part of the solution involved codifying health program knowledge into semantic models (taxonomies and ontologies) for graph ingestion and retrieval. Custom taxonomies categorize end-user responses and enrich survey questions with metadata, enabling efficient retrieval for AI conversational experiences and agentic loops. Developed in collaboration with subject matter experts, these taxonomies include alternative labels and terminology to reflect how researchers search for information.

Leveraging a Semantic Layer for Research Curation and Conversational Experiences

Custom ontologies support data ingestion, classification, and knowledge domain connectivity. They provide a graph schema and structure for data extracted from surveys and transcripts, facilitating insight retrieval. Built using the Resource Description Framework (RDF) and Semantic Web Standards, these ontologies are extensible to other organizational areas and future datasets.

Solution Build

1. Data Enrichment & Graph Curation:

EK developed a reusable pipeline that transforms unstructured surveys and transcripts into a populated graph of connected data. This pipeline uses a large language model (LLM) to automatically classify questions and responses based on the custom taxonomies. The pipeline also leverages the ontology as the schema to build an instantiated graph of connected data. Leveraging the ontology as its schema, the LLM takes tagged statements and inferred relationships to automate the creation of triples required for the RDF knowledge graph.

2. RDF to LPG Transformation:

EK developed a reusable framework to transform RDF graph data into Labeled Property Graph (LPG) data. This transformation provides the final LPG with new semantic context unavailable in out-of-the-box LPGs, unlocking new insights and improved inference for chatbot responses.

3. Chatbot Development

EK designed and implemented an agentic AI solution that translates natural language questions into queries using semantic domain models. This AI agent parses incoming questions based on modeled concepts and user intent to optimize retrieval requests. Once records are retrieved, results are summarized and pulled together in a custom front-end interface developed by EK. An additional AI-agent brings together the records to formulate a chatbot response in the front-end containing both a summary and links to relevant data sources.

Solution Architecture & Roadmap:

To ensure scalability, EK developed a future-state semantic architecture and roadmap to guide how the organization uses semantic tools and unifies multiple AI projects. Leveraging EK's proprietary Semantic Layer Maturity Benchmark, the roadmap aligns the factors for implementing an enterprise-wide Semantic Layer as a Service. Furthermore, the architecture consolidates multiple AI pilots and initiatives occurring across the organization. This new framework provides a framework for how AI agents reuse components and provide insights across various programs.

Leveraging a Semantic Layer for Research Curation and Conversational Experiences

The EK Difference

EK provided specialized expertise in semantic modeling, graph architecture, and technical strategy to successfully deliver both the semantically backed chatbot and the repeatable system architecture. The program's technical components were designed for interoperability, allowing the organization to extend and reuse them across future search and conversational AI use cases.

EK provided end-to-end support to the organization throughout the engagement, leading the program through strategy, design, and implementation phases. This approach ensured that both business and technical requirements remained aligned from the project's inception to its long-term sustainment.

The Results

The semantic layer solution provided the health program with new qualitative research insights, repeatable data transformation processes, and a connected data framework for future solutions. The enrichment pipeline automatically transforms unstructured surveys and transcripts into machine-readable data, making fresh data available to health program researchers. The pilot chatbot saves the program researchers time by generating inferred insights on product experiences, a task that previously required high manual effort. The semantic models established an expandable, connected data framework that is used to link data and relevant research beyond the initial program.

Business Outcomes

- Designed the technology stack so that 90% of its components can be implemented within existing enterprise systems, saving costs on custom development and reducing technical debt.
- Provided health program researchers with a comprehensive and accessible dataset of transcript and survey data.
- Reduced the steps required to gather qualitative data from 5+ to one, decreasing the time spent on manual analysis.
- Developed a 12+ month plan for productionalization and pilot scaling, preparing the organization for future development and broader AI initiatives.

Leveraging a Semantic Layer for Research Curation and Conversational Experiences

As a result of this work, the organization is expanding how it leverages knowledge graphs, semantic models, and AI agents to support researchers and decision-makers. EK laid the groundwork for a new operating model and a semantic solution team that will continue to enable teams and end-users across the enterprise.