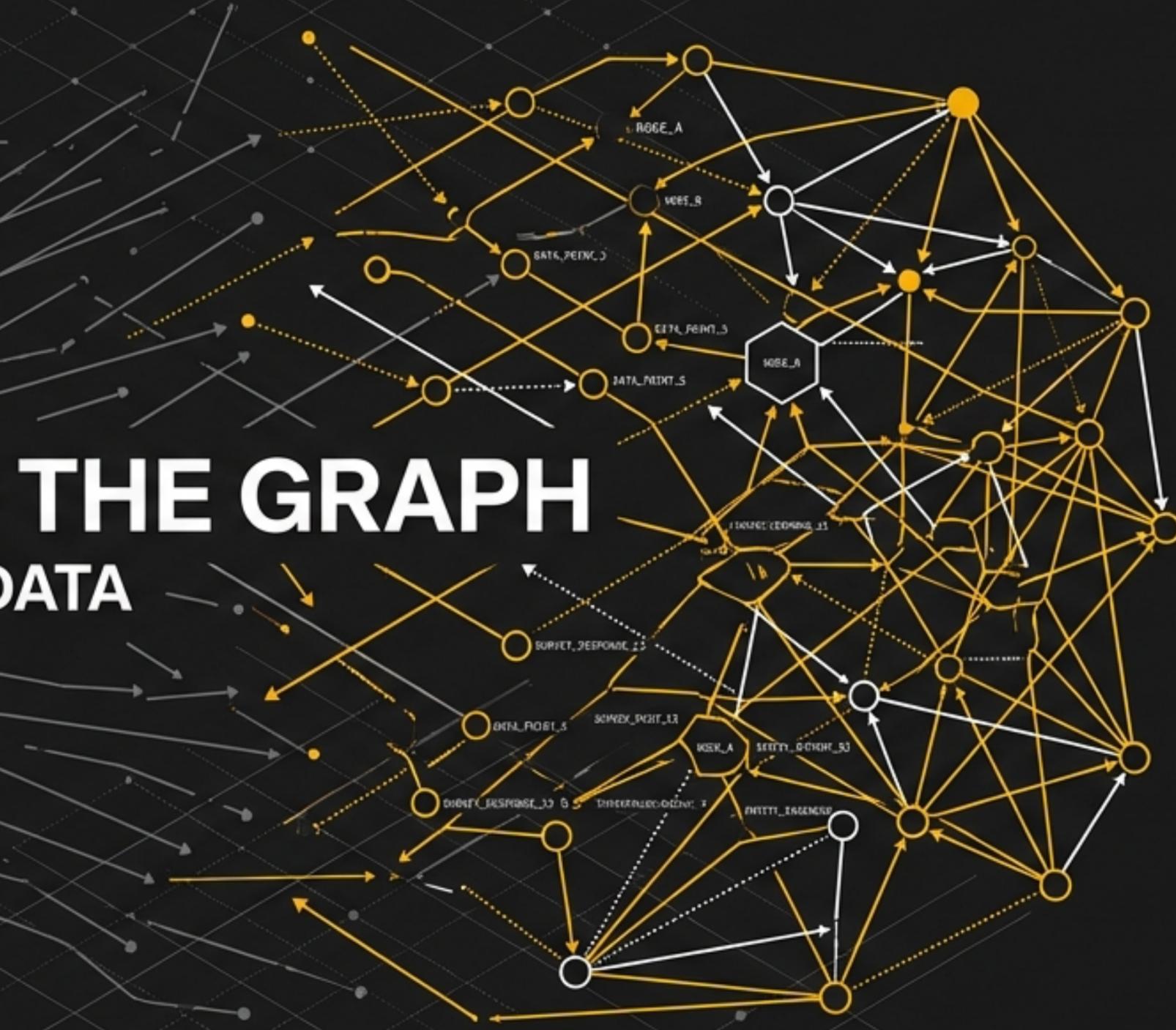


Row 1, Col 1	Row 1, Col 2	Row 1, Col 3	Row 1, Col 4	Row 1, Col 5	Row 1, Col 6
		Data 01	Value X	Timestamp: 1278900000	
...	...	...	...	...	...
		Data 02		10 550	

# FROM FLATLAND TO THE GRAPH

## THE SEMANTIC EVOLUTION OF DATA ENGINEERING SURVEY DATA

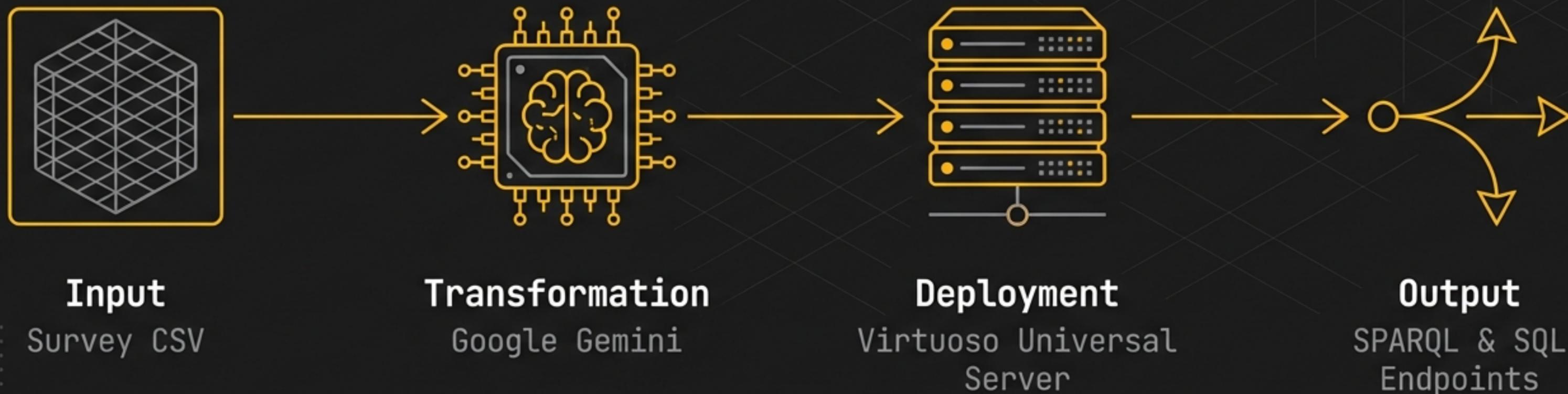
		Timestamp: 1278780000	Value X		10 350
				Timestamp: 1278900000	
...	...	Question 1 Answer	...	Question 1 Answer	...



A journey from 1,100+ CSV rows to a fully interoperable Linked Data ecosystem.

# A New Source Navigation

The objective is to liberate static survey data for sophisticated reasoning. We are taking a raw analysis dataset and transforming it into a Knowledge Graph.



The goal is not just storage, but integration with global knowledge bases.

# The Primitive State: Trapped in Two Dimensions

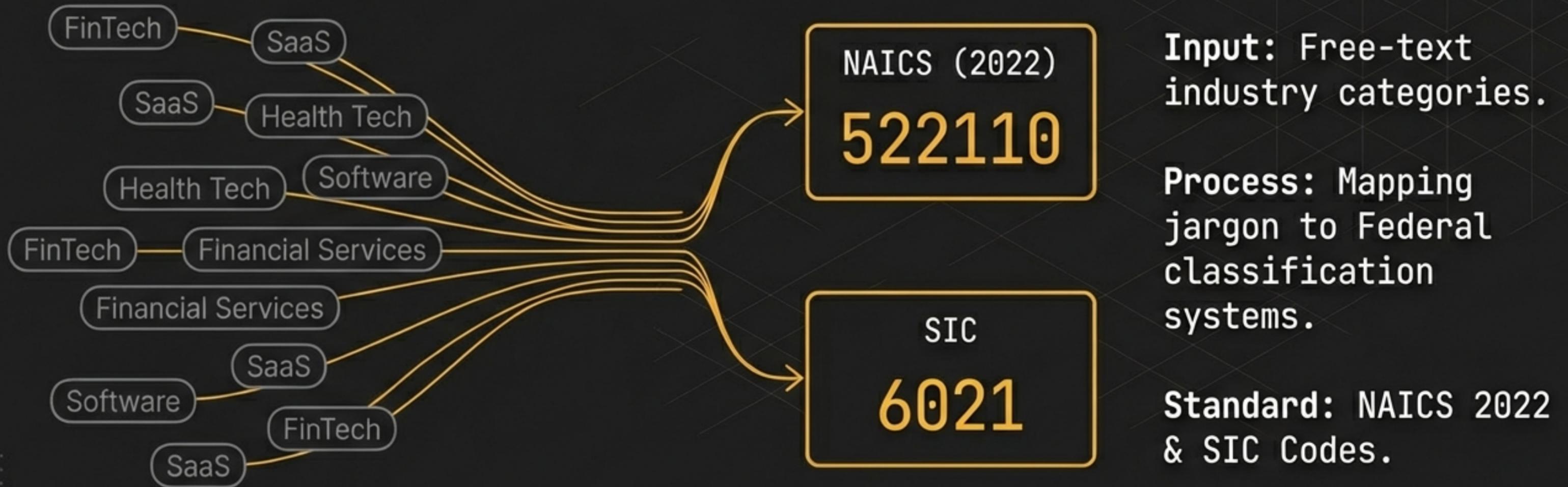
Source: 1,100+ Participant Responses

Context: 2026 Data Engineering State of the Union

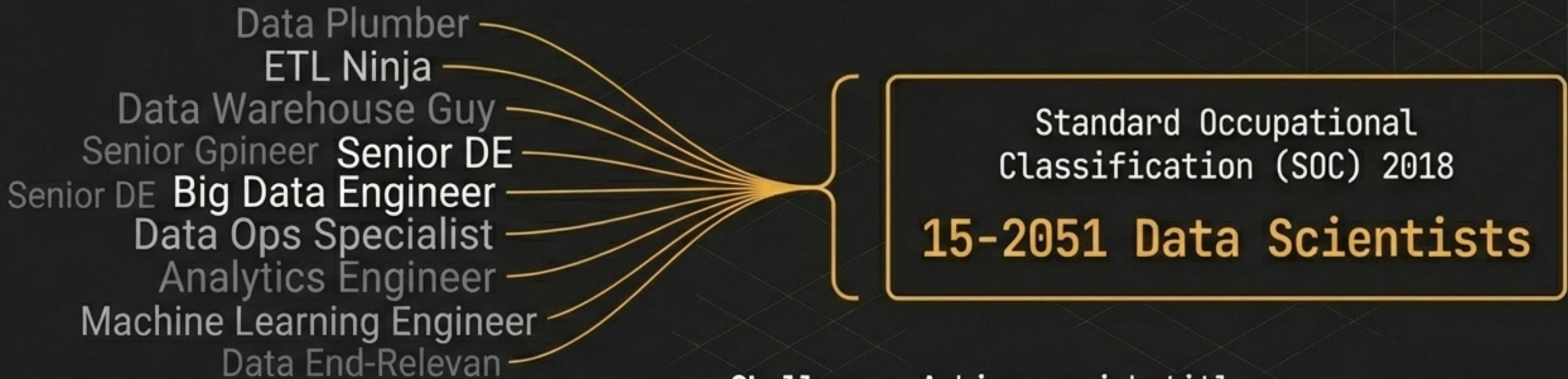
Deficit: Flat architecture. Ambiguous free-text. Zero interoperability.

Valuable insights are locked in a format that prevents reasoning.

# Cognitive Leap I: Industrial Standardization



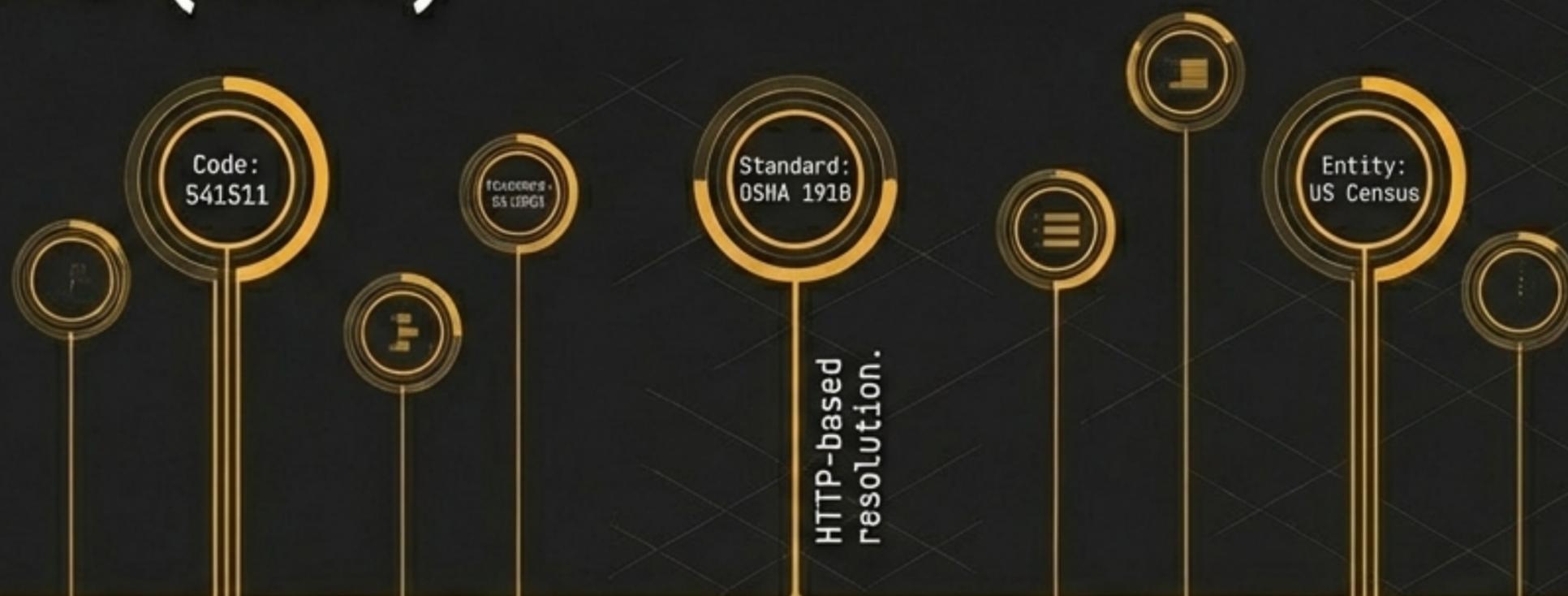
# Cognitive Leap II: Occupational Resolution



- **Challenge:** Ambiguous job titles.
- **Technique:** Keyword analysis and semantic alignment.
- **Result:** Official government standardization.

Valuable insights are locked in a format that prevents reasoning.

# Grounding via Internationalized Resource Identifiers (IRIS)



OSHA



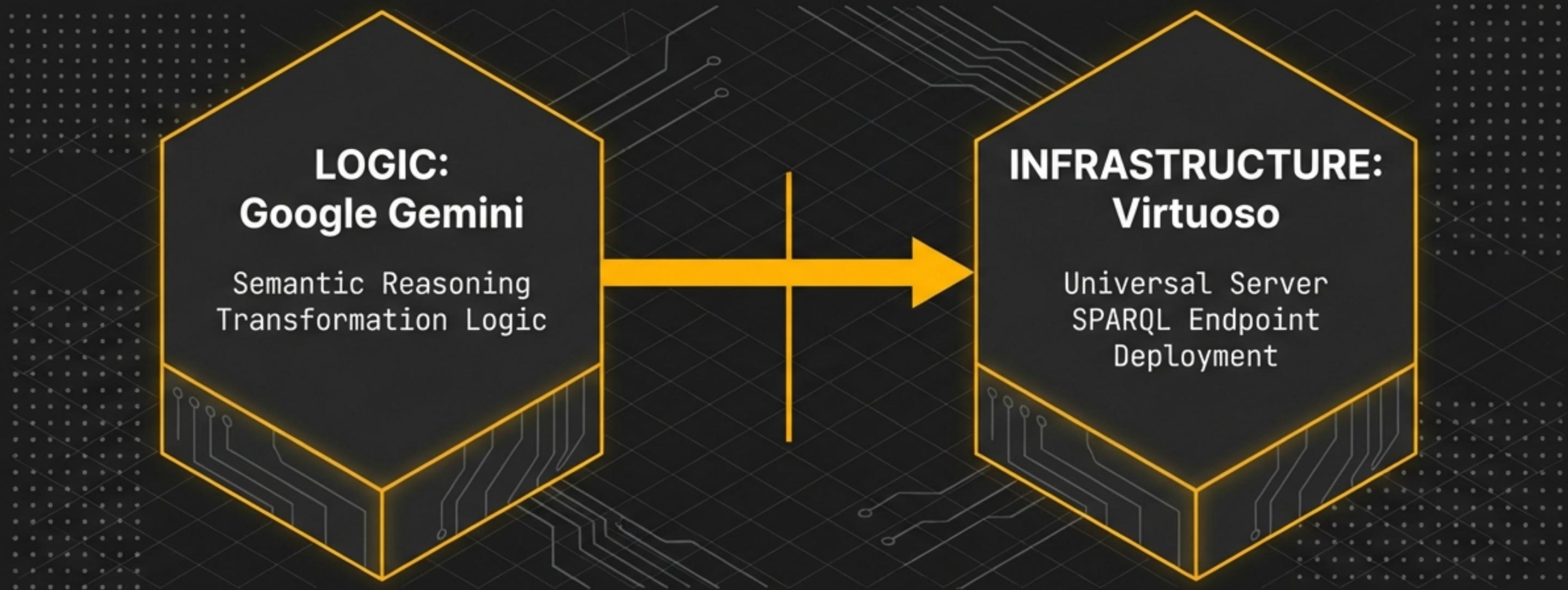
US Census Bureau



O\*NET

“Moving from distinct text strings to resolvable Web entities.”

# The Transformation Engine



AI provides the logic; Virtuoso provides the infrastructure.

# Serialization: RDF-Turtle

Subject (URI)

• `<https://gemini.google.com/share/e37d...>`

`a :SurveyResponse ;`

Predicate

• `:hasIndustry <http://naics.gov/541511> ;`

Object

• `:hasOccupation <http://soc.gov/15-2051> ;`

`:reportingYear "2026"^^xsd:gYear .`

• Format:

**RDF-Turtle (.ttl)**

• Volume: **28,000+**

**Triples Generated**

# Hashing Strategy & Identity

e37d9615f322#survey\_entry\_450



Ensures stable, unique identity for every node in the graph.

# Syntactic Robustness & Literal Preservation

- **Lossless Transition:** Responses stored as literal properties.
- **Triple-Quoted Literals:** Accommodates multi-line text.
- **Strict Escaping:** Handles complex punctuation safely.

```
"""I wish we had better  
documentation for the  
legacy systems! It's a  
nightmare."""
```

# The Ontology: RDFS/OWL Formalization

- **Purpose:** Providing a rigorous schema for the Knowledge Graph. Schema for the Knowledge Graph.
- **Utility:** Defines exactly what each property signifies for machines and researchers.
- **Standards:** Built on RDFS and OWL.



# Governance & Vocabulary Alignment



We are not an island. We are part of the global Linked Data ecosystem.

# The Deliverable Suite



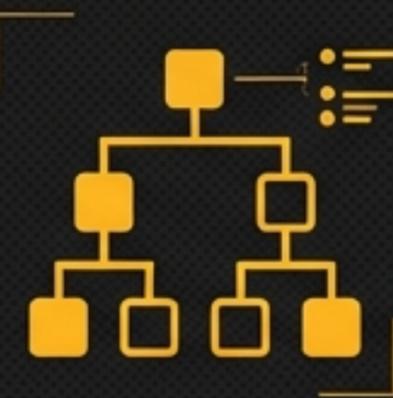
## Enriched Spreadsheet

Hybrid CSV + Federal  
Codes + IRIs



## Knowledge Graph

.ttl File  
(28k+ Triples)



## Formal Ontology

Semantic Roadmap

# Unlocking Advanced Reasoning

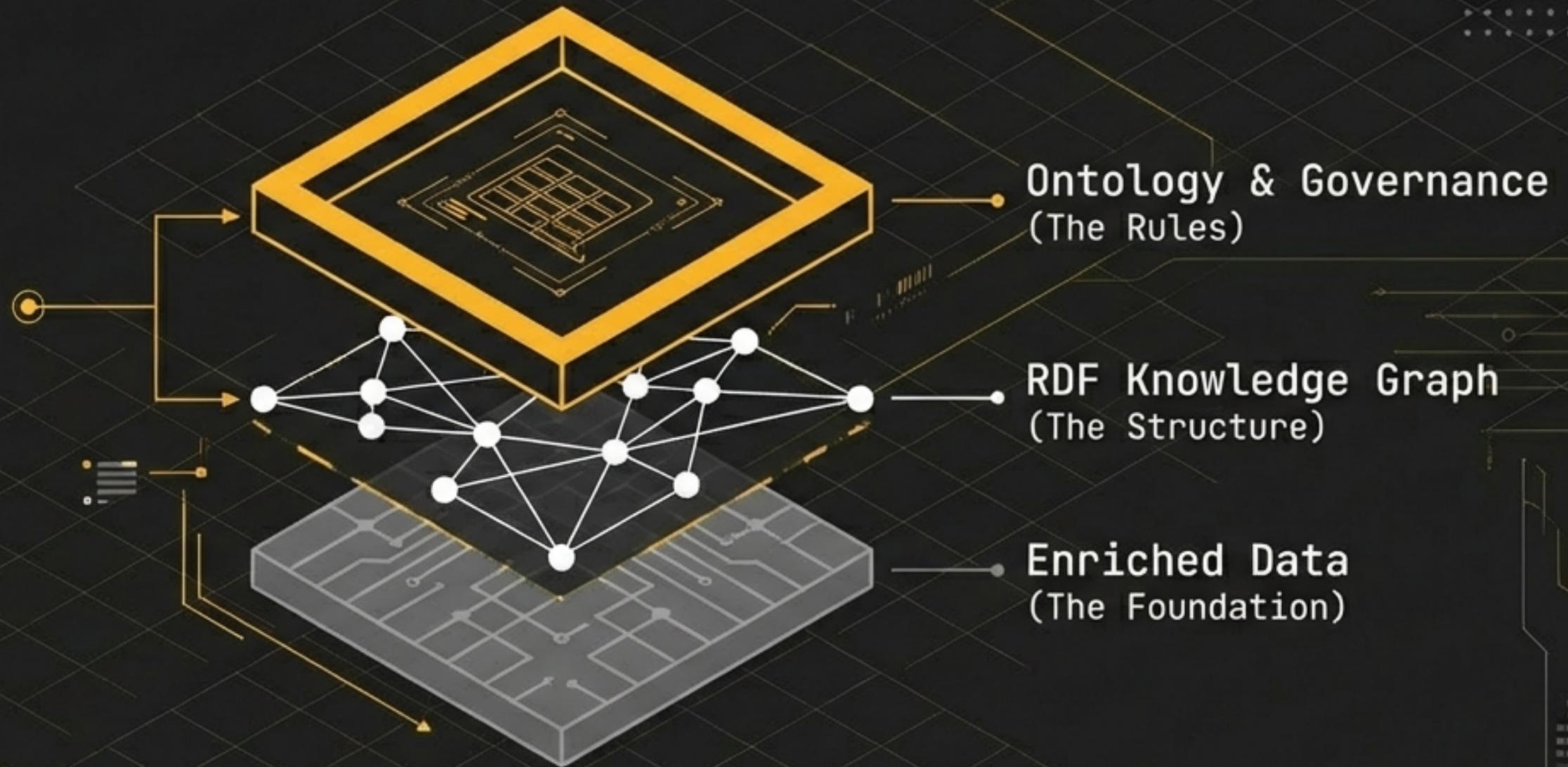
SPARQL  
Analysis  
Ready.

AI Adoption  
Rates

Bottleneck  
Identification

From static  
storage to  
active  
questioning.

# The Complete Semantic Stack



The data is no longer static. It is alive, connected, and ready to answer.