



SOLVING THE DATA INCONGRUENCE DILEMMA

THE ENTERPRISE KNOWLEDGE GRAPH VALUE PROPOSITION

BY MICHAEL ATKIN

The infrastructure for managing data across the financial industry is built on 50-year old technology. Line of business and functional silos are everywhere. They are exacerbated by relational database management systems based on physical data elements that are stored as columns in tables and constrained by the lack of data standards. Enterprise Knowledge Graphs are being implemented to uproot this chaos.

Solving the data silo (integration) challenge is a baseline objective for many organizations. Data meaning is tied to proprietary data models and managed independently. These data silos when combined with external models for glossaries, entity relationship diagrams, databases and metadata repositories lead to incongruent data. Tables have specified formats, fixed numbers of columns and multiple data types. Relationships must be explicitly described and are implemented with foreign keys and multiple joins that are hard to unravel. As organizations grow and applications multiply the sheer number of physical elements in systems continues to increase. The more redundant systems the organization has (or obtains via acquisition) the more bespoke data elements exist



and the more technical debt we accrue. Most reconciliation efforts are based on the hope that we can somehow align these siloed systems with their pre-defined and rigid schemas. The reality is that because of the explosion of uniquely labeled elements, it is nearly impossible to align these silos. Specifications are hard-wired into solutions that use different definitions and column names and that serve a variety of business functions. Modified meaning and incongruence are a symptom of highly siloed systems and expose the limitation of the technology paradigms that currently exists.

Some firms have tried to address this problem by creating a “single version of truth” in the form of one canonical data model for all use cases and all analytical objectives. We have discovered, after many years of effort, that there is simply too much



complexity for every entity and every endpoint to be connected to a single model of data. Complex analysis requires filtered information – combining identity, meaning, time and source to maintain context. It is just not practical to create one data model. It would be hard to construct, out-of-date before deployment and the cost of maintenance and alignment would consume significant resources.

The lesson from this effort is that unconnected data is a serious liability. Data that is hard to access, blend, analyze and use impedes application development, data science, analytics, process automation, reporting and compliance. Requirements-led design in silos and the goal of a “one size fits all” approach is what created the federated data problem in the first place and results in more semantically misaligned silos with little possibility for leverage or reuse.

Pathway to Solution

The problem of data silos and fragmentation is reversible by applying web standards and semantic technology in the form of an enterprise knowledge graph (EKG). Enterprise knowledge graph technology provides an expressive model that is both conceptual and operational. It is built on data standards that can be reused. And it has the capacity for inference and reasoning that expedite the migration to AI and machine learning.

Semantic technology is the best way to handle more data that is coming faster and in a wider variety of forms. It was designed specifically for interconnected data and gives us the ability to unravel complex relationships. This shift from conventional technology to EKG technology results in our ability to describe what the data means as

well as how concepts are connected. It eliminates the need to reconcile foreign keys and join tables. By unshackling technology from hard-coded schemas and rigid data structures we can be more agile in our operations and more flexible in our analytics.

Concept Reuse

The advent of semantic technology creates a more powerful user environment and enables business users to work with concepts (which they intrinsically understand) without forcing them to work with the underlying physical elements. Semantic modeling eliminates the problem of hard-coded assumptions because it focuses on concepts, not specific applications. Users always understand what the data represents even when it moves across organizational boundaries. This enables efficient data reuse across systems and processes and allows multiple models of the same data to coexist. The semantic approach allows firms to map data once but leverage it many times.

Instead of data silos we get data that is integrated and linked. Our organizations become more efficient because ontologies are standardized and reusable. With semantics we get economies of scale because we don't have to continually reinvent the wheel for identical concepts. Reusability gives us significantly lower costs and faster time-to-knowledge.

Data Integration

The capability of the enterprise knowledge graph to resolve identity, meaning, time and source are the keys to the harmonization of data across our organizations. The core of the value proposition starts with the ability to connect data elements to a universally unique, web-addressable identifier. The



identifier enables firms to link data wherever it resides, eliminating the need to continually map data across the enterprise. The web-based ID becomes the Rosetta stone for identity resolution.

The EKG uses conceptual models (ontologies) to precisely describe what the data means as well as how concepts are connected. These ontologies are used to align business glossaries and can be directly translated into physical data structures. The EKG establishes shared meaning across fragmented sources of data. It is flexible and can traverse connections to identify complex relationships. Linking data to meaning rather than to applications enables users to analyze data from many perspectives.

The identity and meaning resolution capabilities of semantic technology enables firms to track data at its most granular level. This provides assurance that the firm is getting data from the right authoritative sources and allows us to recreate the value of the element when the data was created. As a result, lineage and provenance are traced across complex data systems. The EKG becomes the logical distribution point because it traces data flow (and is fully auditable) by source, purpose and responsible party.

Data Quality

Enterprise knowledge graphs use ontologies and semantic constraint rules to identify violations of logic and definitional conflicts. Data quality and structural business rules are linked to the logic of these rules is captured and expressed as executable models and consistently enforced across all systems and processes. These quality constraint rules enable firms to measure quality and perform automatic data validation across systems. Any violation of logic

is identified and prevented before data enters the system. Rules can be modeled for all circumstances and controlled at both the applications and data level. This means that security is embedded into the design of the data and not constrained by either systems or administrative complexity.

Conclusion

This article was structured to shine the light on the limitations of conventional technology to solve the data harmonization needs of complex organizations. My objective was to demonstrate why semantic technology is the more efficient and productive way forward. Think of it as the content infrastructure to a new world of facts and relationships about people, processes, applications and data.

But addressing data incongruence is only half of the story. Enterprise knowledge graphs also stimulate cross-departmental and interdisciplinary communication. They help to orchestrate information flows and unravel complex data relationships. Enterprise knowledge graph is a prerequisite for achieving smart, semantic AI-powered applications that can help you discover facts from your content, data and organizational knowledge which otherwise would go unnoticed. Enterprise knowledge graphs help you organize the information from disparate data sources to facilitate intelligent search. This is a game changer that leverages natural language processing, semantic understanding and machine learning as part of the knowledge graph environment. Data becomes understandable in business terms and is no longer obscured by technical definitions that are meaningful to only a handful of specialized personnel.

And finally, knowledge graphs spur digital



transformation by delivering a “digital twin” of your organization that encompasses all data points as well as the relationships between data elements. By fundamentally understanding the way all data relates throughout the organization, the enterprise knowledge graph offers an added dimension of context which informs everything from initial data discovery to flexible analytics.

We are now standing at the new precipice of content interoperability. This is one of the most important developments in terms of productivity and will be a revolution for the management of knowledge. It is time to make the leap.



Michael Atkin is a principal at agnos.ai and Director, Enterprise Knowledge Graph Foundation. He is a financial industry analyst, founder of the EDM Council, and lecturer on the principles of data management at Columbia University. *He can be reached at Mike.Atkin@ekgf.org*

The Enterprise Knowledge Graph Foundation (EKGF) is a non-profit organization focused on growth and development of the marketplace for semantic standards. The EKG Foundation was established around three fundamental objectives: (1) position knowledge graph technology as the most effective way to connect data across the enterprise, (2) promote best practices for EKG implementation, and (3) advance industry collaboration through a number of portals that will accelerate adoption. For more information, visit EKGF.org.