

FINANCIAL SERVICES

Leveraging Semantic and Graph Technology to Tame the Enterprise Data Storm



Introduction

Financial services companies are under great pressure to minimize risk, cut costs and boost competitiveness amid a perfect storm of rapid legal, business and technological change.

Organizations must proactively modernize their regulatory, compliance and surveillance programs so they provide genuine value beyond merely "checking the boxes". They must also pursue new revenue opportunities made possible through advanced technological initiatives such as customer 360.

Achieving these vital goals requires effectively leveraging more data drawn from more disparate data sources than ever before; however, traditional data management solutions are no longer adequate to fulfill these challenges.

New, modern data discovery and integration technology is required to enable essential new and deeper levels of analytic insight, beyond what is possible using legacy data solutions.

This white paper describes the rapidly evolving business and regulatory drivers leading financial services organizations to implement vital new enterprise-wide data technology initiatives, and why legacy data management tools can no longer keep up.

You will also discover a new, modern approach to enterprise data discovery and integration, powered by semantic and graph technology, capable of integrating, blending and presenting disparate data sources – structured and unstructured, internal and external alike. This paper concludes with a success story in which Anzo, a proven semantic and graph-driven data management platform, made an enterprise data integration strategic initiative an exciting new reality for a leading financial services organization.



Key Challenges

Compliance with a wide array of regulations and high priority risk management and surveillance initiatives are consuming IT budgets, eroding gross margins, and impacting resources for new revenue generating projects.

Regulatory compliance, risk management, surveillance, competitive pressures, and the mandate to reduce operating expenditures in today's new lower margin business environment are forcing financial services organizations to fundamentally transform their ability to manage, utilize, and monetize their disparate and inconsistent enterprise-wide data assets. Doing so requires the ability to access and analyze disparate data sources in a unified, flexible manner at massive scale.

Consider the challenges associated with insider trading activity monitoring. Limiting surveillance analytics to activities related to a particular security or an employee's trades is clearly insufficient. Building a comprehensive picture of potential disclosures of non-public information requires linking and exploring a wide variety of data sources and types; including:



Manual navigation and exploration of relationships spanning these and other data sources requires substantial time and

resources, leading to surveillance efforts that are limited in scope, minimally effective and operationally expensive.

Recognizing the critical need for a new and deeper level of enterprise data integration and insight, financial services firms are launching new enterprise data initiatives that cut across traditional data silos based on line of business, geographical location and data sources. These initiatives require close alignment among organizational structure and oversight, a common enterprise *lingua franca* (business glossary) and new data management technology that eliminates the architectural shortcomings that have resulted in the current fragmented data landscape (*see figure below*).



A common lingua franca defines a representation of enterprise knowledge that can be used to drive the harmonization of otherwise disjointed and ambiguous enterprise data. For example, the EDM (Enterprise Data Management) Council is leading a large diverse consortium of members to create FIBO (Financial Industry Business Ontology), an emerging lingua franca for the financial industry sector developed by consensus of a broad membership of contributing subject matter experts. The resulting FIBO master can provide a valuable and extensible



financial business glossary that can be leveraged as a starting point for developing a common lingua franca.

Unfortunately, the implementation of an effective technology architecture to support comprehensive enterprise-wide data initiatives has been extremely difficult and largely unsuccessful to date. This is due in part to the large existing fragmented and heterogeneous enterprise data landscapes that have evolved over decades, resulting from bottom up implementations using different systems with different architectures, data types and formats, without a standard lingua franca.

Additionally, traditional data management technologies and newer big data technologies do not easily lend themselves to support these kinds of enterprise data harmonization initiatives. Traditional relational database technologies including data warehousing, multiple SQL queries and procedural data ETL processes have contributed significantly to the problem of inflexible data silos and are not the answer.

Big data frameworks like Hadoop can provide cost-effective repositories for large disparate data sets and simple search capabilities; however, they do not provide the unified common data architecture or the ability to flexibly navigate and leverage dissimilar data from disparate sources.

What is needed to successfully navigate the many challenges facing financial services firms today is an entirely new approach to enterprise data management that can be seamlessly incorporated into existing heterogeneous data environments to harmonize all of the internal and external data to meet regulatory, operational and cybersecurity pressures, and capitalize on new business.

Semantic & Graph Technology: Modern Data Management

Semantic and graph-driven enterprise data discovery and integration empowers organizations to meet regulatory, operational and cybersecurity pressures and capitalize on new business opportunities in ways beyond what is possible using traditional relational data management solutions.

Semantic and graph-driven data management platforms enable financial services organizations to:

- Freely combine, harmonize and blend existing and future enterprise data assets – structured and unstructured, internal and external – in a repeatable, future-proof manner
- Provide flexible, interactive ad hoc access to customized data using all of the data to meet the demands of business users, industry regulators, and analysts

Semantic and graph data technology provides modern data modeling, data integration, knowledge representation and on demand access to analytics-ready data. to o the new deeper level of data integration and on demand data access demonstrable control over disconnected and inconsistent data, regardless of data type or source.

It uniquely supports unplanned ad hoc queries and data exploration by business users and regulators without requiring complicated and time-consuming pre-processing of the data in advance to support specific downstream queries. And it facilitates the application and enforcement of a business glossary as a standard lingua franca spanning the underlying disparate and inconsistent raw data assets, enabling the creation of an enterprise-wide business data catalog that can be leveraged and enforced throughout the entire organization, while preserving the fidelity of the original source data, maintaining data lineage, and providing an execution point for applying proper data governance.

Semantic & Graph-Driven Modern Enterprise Data Management: Key Benefits



Simplifies access to complex and blended data to address unanticipated questions



Quickly profiles, connects and harmonizes structured and unstructured data alike from multiple sources



Presents customized data views based on user roles with conceptual models using business terms

Flexibly accommodates new data sources and use cases on the fly with minimal impact



Scales horizontally to accommodate massive enterprise scale

An important aspect of semantic technology is that it focuses on the meaning of data by formally representing the relationships between data assets as well as the data assets themselves in a schema-agnostic fashion. Data is represented atomically and can be used to construct any information mart required to support the full breadth of enterprise query types and complexities ranging from relational to graphical (see figure at top of next page). A **semantic layer** is a human-readable data representation that **uniquely identifies and connects data with common business terms** that helps users access customized, analytics-ready data.



Using such a semantics approach to represent any data asset and its relationships to other data assets enables the creation of a connected knowledge graph capable of spanning organizational context (*see example below*).







This harmonized linked data approach supports an exciting new level of data discovery; one in which business users can flexibly navigate the data and relationships, iteratively asking questions and getting answers on the fly, allowing them to ask derivative questions based on answers to their questions, taking them into unplanned directions. This type of iterative ad hoc data exploration and discovery is difficult or impossible with traditional relational technologies or with disjoint, ambiguous data lakes, which require significant expertise in all the relevant source data schemas to properly construct highly complex caseby-case aggregating queries and require cumbersome and timeconsuming data preparation tasks.

Semantically harmonized data spanning large numbers of disparate data sources abstracts the query user from the complexity of linking disjoint data schema, resolving ambiguous identities and trying to reverse engineer the intended meaning of primary key - foreign key relationships on a query-by-query basis. Instead, this work is done once by applying a business glossary upfront, removing the need for complex query-byquery data janitorial work. The end user can navigate data quickly and easily, leveraging common semantics that include formally expressed relationships.

When operating at big data scale (e.g. petabytes of data), semantic technology can provide financial services organizations with capabilities it desperately needs to apply a comprehensive business glossary to its enterprise-wide data assets, and to navigate and leverage the data for risk management, compliance, surveillance, and to exploit opportunities for growth in ways that are simply unattainable with other technologies. Even using semantic technology, however, it can be impractical for organizations to attempt to completely implement a technology architecture in an overarching top down manner. These kinds of top down



enterprise data initiatives are not usually successful. A much more pragmatic approach is to implement specific projects on an incremental basis at the departmental level, aligned with enterprise-wide organizational governance, delivering quantifiable value with each implementation and enabling seamless re-use of the underlying data assets for future initiatives. Semantic technology is well suited to support this type of incremental approach to implementation.

A set of use cases for which semantic technology is ideal is regulatory compliance, since it often requires leveraging a diverse set of structured and unstructured data from many different sources (e.g., emails, contracts, applications, databases, etc.). For example, applying semantic technology as part of BCBS 239 compliance provides an execution framework for concretely applying the business glossary to existing and disconnected enterprise data, providing a repeatable way of generating consistent and connected high fidelity data from multiple authoritative sources for regulatory reporting and analytical models. Such a framework can generate a further ROI by enabling the same data assets as high-fidelity signals into broader revenue generating initiatives such as 360-degree view of the customer, client targeting, and enhanced operational and market intelligence.

The ability to seamlessly re-use the harmonized data assets from prior implementations for new initiatives enables organizations to more easily implement projects that increase revenue and margins, even amidst an explosion in regulations, an increase in the volume and complexity of their data assets, and strained IT budgets.

Anzo®: A Proven Semantic & Graph-Driven Data Platform

ANZO

Anzo[®] by Cambridge Semantics provides proven semantic and graph-driven data management, discovery and integration platform, leveraging semantic and graph technology at its core.

Anzo is a modern data discovery and integration platform that lets anyone find, connect and blend any enterprise data into analytics-ready datasets.



Anzo works with any combination of diverse structured and unstructured data to enable organizations to manage, leverage, and monetize their disparate and inconsistent enterprise-wide data assets. Anzo enables organizations to gain new and deeper levels of vital insights on demand at enterprise speed and scale.

The functional capabilities of Anzo can be described in terms of three key areas:

- Business-to-data modeling
- Accelerated insights at scale
- Enterprise grade capabilities



Business-to-data modeling. Anzo uniquely addresses the key data challenge in large financial services organizations: how to gain control of the vast, disparate data across the enterprise, often defined and managed at the line of business or application level, with different representations, terminology, formats, and so on.

Anzo enables organizations to visually define a common business glossary across the entire organization that bridges the business and all of its data assets. Users can define business glossaries and import and extend standard ontologies provided by third parties or industry groups, such as the Financial Industry Business Ontology (FIBO, a business conceptual ontology developed by the members of the EDM Council). Anzo supports both top down and bottom up implementation approaches. Organizations can begin by cataloging information assets, or they can categorize them on the fly as assets are leveraged for specific value-generating data driven projects, simplifying implementations and generating faster ROI.

Anzo provides capabilities to easily extract data from sources, and link and map data and structures to the business glossary to create enterprise data catalogs. Anzo ingests both structured and unstructured data with extremely high performance at scale, leveraging horizontally-scaled processing techniques, mapping and transforming data from all sources into semantic models. Ingestion of most structured data is automatic and can eliminate the requirement for many manual mapping steps. User defined mappings and links, transformations, and event driven rules can be defined through an intuitive business user interface similar to Microsoft Excel, allowing business users to create the catalogs. Anzo also provides a power- user workbench for more technical users. Mappings to source systems and linkages between data sets are created with provenance for trust and

traceability. Access control, security, full data lineage or provenance, and data context can be defined and preserved.

The ability to include unstructured text data can provide invaluable insights for supporting risk management, surveillance, and compliance initiatives like AML, KYC, fraud detection, and so on by allowing users to explore the text and meaning in documents, emails, instant messages, and other text sources in a completely integrated and seamless manner with all of the enterprise data. Including unstructured data in the analyses is also important for many revenue generation initiatives; for example, to include customer emails and notes from free text fields in CRM systems for 360-degree view of the customer initiatives.



Accelerated insights at scale. The data-to-business modeling in Anzo results in an enterprise-wide knowledge graph that can be flexibly explored and navigated without involvement from IT, and without requiring knowledge of the details and structure of the underlying data. Business users, analysts, researchers, and regulators can ask questions of the entire corpus of enterprise data. To support the way that users think, billions of facts are available to be simultaneously queried interactively.



Such large-scale computing is enabled by AnzoGraph® DB, the analytical graph database developed by Cambridge Semantics, embedded within the Anzo platform. AnzoGraph is a massively parallel processing, in-memory analytics engine that ingests large volumes of graph data into memory and can execute large numbers of interactive requests that resolve as complex ad hoc "join" queries and filters. This ability to support users' needs to explore, pivot, and combine (e.g. join) semantic data at enterprise scale with the highest level of performance is a requirement for successful enterprise-wide data initiatives, and this feature is unique to Anzo. In fact, Anzo is proven to be more than 100 times faster at scale than any other vendor.

Anzograph DB holds the performance record for the Lehigh University Benchmark (LUBM), loading and querying one trillion triples on the Google Cloud Platform more than 100 times faster than the previous record (1.98 hours vs. 200 hours)¹. This level of performance at scale is a critical capability for enabling large financial services organizations with tens of thousands of databases and petabytes of data to leverage semantic technology for their enterprise data initiatives. It is also a requirement for enabling regulators to get answers to unexpected questions in near real time, without having to involve IT or re-structure the data, eliminating penalties and lowering capital requirements for the firm.

Anzo also includes fully integrated visualization, reporting, and dashboarding capabilities that allow business users to visually explore and navigate the data and relationships in a unified, holistic, and consistent manner without requiring any type of specialized query or syntactic knowledge. Users can visually and interactively traverse even the most complicated multidimensional data, helping them glean knowledge from the data and easily pivot to take data exploration in new directions.

1 Trillion-Triples Benchmarking White Paper, Cambridge Semantics, Inc., December 2016.



Users can drill into any of the data using the visualization tools to introspect any aspects of the original source data, and for tracing lineage and provenance, for example to identify the source of the data, how and when it was transformed, and by whom. With IT-controlled permissions, researchers can use dashboards and analytics created by others so that efforts of one user can be leveraged as the starting point for another user's discovery and analytics.

In addition to providing its own native visualization capabilities, Anzo also provides integrations with external reporting tools like Qlik and Tableau to support existing dashboards and reporting initiatives already in place, allowing users to view and explore the underlying data in Anzo from their existing, familiar reporting tools. Business users can continue to use the reporting tools that they are familiar with, augmenting them with the ability to seamlessly navigate and explore the underlying enterprise data models developed using Anzo.

Enterprise grade capabilities. Anzo provides the enterprise level access control, security, governance, lineage, and versioning capabilities required by major financial services organizations. It is built using open standards, eliminating the risk of vendor lock-in. It supports both on premise and cloud deployment and can be scaled out and in on-demand in response to processing loads, minimizing operational costs.

For example, if 100 or so on-demand cloud "graph marts" are needed for a processing task, the organization can scale the cluster out, then back in, using commercially available cloud providers. Anzo provides high availability and failover capabilities to support mission critical applications and to minimize unplanned downtime. Anzo supports integration with LDAP directories and other security systems to provide enterprise identity and access control.



Anzo in Action: Success Story

State Street Bank, EDM Council, dun & bradstreet, and Wells Fargo collaboratively utilized the Anzo platform to harmonize State Street's Interest Rate Swap data with dun & bradstreet's entity hierarchy data, using the Financial Industry Business Ontology (FIBO).

State Street Bank defined the business requirements and supplied operational data, the EDM Council supplied the FIBO model and recommended reports and analytics, and dun & bradstreet supplied business entity and corporate hierarchy data. Wells Fargo provided FIBO consultation for this project (see figure below).



The Financial Industry Business Ontology provided an excellent starting point for this implementation. Rather than creating a new business glossary and model from scratch, the team was able to leverage parts of the FIBO master. FIBO is expressed in



the standard W3C semantic modeling language Web Ontology Language (OWL), natively supported by Anzo. The power of semantic models like FIBO is that they describe data in business terms and do not limit the questions end users can ask.

Loading FIBO into Anzo was accomplished using a simple import function. There was an excellent match between the FIBO model and the data sources from State Street and dun & bradstreet, and this alignment made the mapping fast and easy. Once mapped, data loading and transformation was automatic. By mapping and loading the diverse data sources using a common model, the data harmonization was automatic. The combined data sets became immediately available for user analytics.

Using FIBO and Anzo, the collaborating organizations were able to rapidly harmonize disparate data sources for conventional analytics as well as more advanced visualization, classification and risk aggregation (see sample analytic view below, showing a network navigation view of companies that are buying and selling interest rate swaps).





Conclusion

Implementing a comprehensive enterprise-wide data initiative is now a top priority for many financial organizations. A successful initiative requires effective coordination among a governing organizational structure, a common business model, and a supporting technology architecture. Cambridge Semantics' Anzo is an end-to-end, fully integrated, semantic data management, discovery and analytics platform for implementing the technology architecture to support these enterprise-wide data initiatives. It can be seamlessly incorporated into existing heterogeneous data environments to harmonize all of the internal and external data, and it is the first and only semantic technology with the performance at scale required by large financial services organizations with tens of thousands of databases and petabytes of data.

By incorporating Anzo as an integral component of an enterprise-wide data initiative, financial organizations can finally begin to gain control of their enterprise-wide data assets to support the demands of regulators, meet operational requirements, and deliver new value-generating services to the business.



About the Authors

Carl Reed was formerly a technology managing director of Credit Suisse, serving as the head of data architecture and semantic technology. Prior to joining Credit Suisse, Reed worked at Goldman Sachs for 21 years where he served in a number of senior divisional CTO and technology strategy roles.

Marty Loughlin is Senior Vice President, Head of Global Sales at Cambridge Semantics Inc. Prior to joining Cambridge Semantics, Marty was the Managing Director for EMC's consulting business in Boston. His 25-year career has focused on helping clients leverage transformative technologies to drive business results, most recently in Cloud and Big Data.

About Cambridge Semantics

Cambridge Semantics Inc., The Smart Data Company[®], is a modern data management and enterprise analytics software company that enables seamless access, integration and analysis of all enterprise data via a graph-driven data fabric architecture.

The company delivers products and solutions that enable IT departments and business users across Life Sciences, Financial Services, Government, Manufacturing, and other industries to accelerate data delivery and to provide meaningful insights across the organization at hyper-speed and scale.

We invite you to explore Cambridge Semantics solutions and discover how a data fabric enhanced by semantic and graph technologies can help your organization unleash the power of your data for digital transformation and competitive advantage. For more information, call +1 617-245-0517, or visit <u>www.cambridgesemantics.com</u>.