



Overcoming insurance data challenges with a semantic data hub



Traditional insurance organizations must adapt to compete in the digital age. In today's evolving insurance market being stuck with complex, inefficient systems risks disrupting the customer experience, hampering revenue opportunities, and magnifying operational costs across lines of business. Insurance organizations need a more efficient, agile approach to enable integration of data across their business.

This paper explores a modern approach to data challenges, and discusses how a semantic data hub enables insurance organizations to improve business intelligence and build smarter applications.

## Data challenges in the insurance industry

With the advent of game-changing technologies, every industry is in a position where existing organizations must adapt to technological advances to compete with new market entrants. Until recently, technology was not able to address the unique use cases presented by the insurance industry, such as underwriting and claims automation. Organizations were forced to shoehorn business requirements into traditional database systems and technological processes unsuited to the diversity of information and data sources of the insurance world.

Newer, point solution InsurTech companies create data silos that are even more impenetrable. A specific subset of customer activity is closed up in a single organization with no interconnects to other data.

Finally, modern technology has advanced to the point where it can address the unique challenges of the insurance industry. As an example, Machine Learning (ML) techniques such as Optical Character Recognition (OCR) can be used to extract data from paper documents. However, digitizing information from documents using OCR alone is not sufficient for automating business processes. This information needs to be linked to industry terms so the right rules can be applied to the scanned information.

Information from sources such as claim forms, insurance contracts, and know-your-customer (KYC) records combines to create a collection of interlinked entities. Their relationship with business terms is a powerful way to build a knowledge base for the application of business rules to specific business processes.

This new approach to intelligent data processing and management is well-suited to tackling technology-resistant problems in the insurance domain, such as:

- Underwriting – armies of underwriters still process hundreds of documents and emails manually.
- Claims automation – a lack of industry-standard business ontology produces mismatched forms, missing information, and confusing cross-communication.
- Customer engagement - insurers need to build a single, consistent – and persistent – relationship with a customer who engages across multiple channels.

To realize efficiencies and capabilities, the modern insurance organization must integrate diverse data and draw connections at an unprecedented scale. Semantic data hubs now allow insurers to pull together all data, documents, and information, extracting meaning and linking it with the context of their business domain.

## Introducing the semantic data hub

What is a semantic data hub, and what is so special about how it handles data? The semantic data hub is built on top of an operational data hub. It establishes semantic relationships between data. An operational data hub is an enterprise architecture pattern built on a multi-model NoSQL database that eliminates data silos and offers a better alternative to data warehouses. Semantics are a special way of building connections between records using a





sophisticated graph database. Together, these two technologies combine to create a semantic data hub.

### Competing in the digital age

To compete in the digital age, traditional insurance organizations must overcome the challenges associated with complex, inefficient legacy IT systems and processes. Key business processes like managing risk, claims processing, and underwriting can be done faster and with more accuracy when the data can be integrated from individual silos to create a unified view of the business.

From a business standpoint, customer experience is the frontier of client acquisition and retention. More effective use of data assets, including semantics, can lead to improvements in customer engagement and allow insurers to keep pace with new market entrants. InsurTech is changing customer expectations, but incumbent insurance companies are well positioned to respond – if they are willing to modernize legacy data systems and processes.

### Using semantics to improve customer experience

A Semantic Data Hub helps develop data-driven applications across processes that positively affect customer engagement.

The creation of knowledge graphs is a key ingredient. Building a knowledge graph requires the extraction of both structured and unstructured data, such as data found in documents and images, and linkage to business ontologies. For instance, MarkLogic-powered Semantic Data Hub from Wipro was deployed to accelerate KYC processes in an insurance company. The new process leveraged a natural language processing engine to parse customer KYC documents, map entities in the document mapped to business entities (e.g., license number, photo identification) and validate data against business rules (e.g. driving license expiry date). This new data-driven process has reduced customer onboarding time from days to minutes.

In another instance, Wipro's Semantic Data Hub was deployed at an insurance carrier with the

objective of increasing the speed of claims processing through semantics-driven machine learning algorithms. Using a smart phone or a web application, the customer uploads vehicle insurance claims with pictures. Behind the scenes a machine learning algorithm processes images, identifies parts which are damaged, and links this information to the car ontology. Next, the system finds the parts, applies the claims rules for the given part, checks the pricing, predicts the repair costs, and creates a permissible claims report. All of this is done in a matter of minutes, compared to the manual, off-line process that took days and sometimes weeks.

Another application of the semantic data hub for an insurance company was for building an ontology-centric pricing application. In an insurance company, actuaries define how products can be configured, how they are priced and at which market segments they are aimed. They document these details as reports and spreadsheets. The introduction of a new insurance product or changes to existing acceptance rules or other aspects of existing products can take months and is subject to interpretation errors. To meet this challenge, the insurance company externalized the business logic of the insurance products into an ontology. The ontology drives the behavior of an application which implements the products definitions. The application efficiently generates quotes, calculates the pricing of contracts and determines the applicability of a product based on an acceptance policy. Changes to insurance products are made by changing the ontology and rules, which, after testing, are used directly by the production application. This has dramatically reduced the time it takes for that insurer to originate new and updated policies for existing and prospective customers.

As these cases demonstrate, semantics enables a whole new class of intelligent insurance applications. Insurers can make new connections, form faster insights, and build more compelling applications for underwriting, claims processing, risk assessment, and customer experience.

## Critical features of enterprise-class database systems

An enterprise-class database system needs to provide the capabilities and features required to compete in a fast-moving market. The best of these systems is uniquely architected to support creation of more intelligent applications.

### Smart mastering

Insurance databases need to deal with incoming data from a wide range of sources. A master record is the result of harmonizing and merging multiple disparate records to create a single source of truth. Smart mastering maintains a history of all merged data and the automated and manual decisions made during harmonization in order to provide the most comprehensive view of a single master record.

This ensures that a smart master record maintains the best possible version of all available data from all merged sources, cutting down on discrepancies and user error while increasing the trustworthiness of data. With a comprehensive data history, it also allows records to be safely and accurately un-merged if the need arises.

### Semantics

Traditional relational databases store simple, machine-friendly relationships between records. But insurance is a human industry, and the relationships between insurance records need to reflect human understanding.

A semantic database addresses this problem by storing relationships that are derived from subject-predicate-object constructions. For example, John (subject), is the main driver of (predicate), car with registration number ABC123 (object). This method of storing simple facts allows records to be linked together to form a network of hundreds of billions of facts and relationships. This web of relationships is key to building a 360-degree view of an entity such as a customer, broker, or property.



## Security

With the complexities around industry and data privacy regulations, insurance companies need to leverage their data in a governed matter. This is practically impossible with most databases, which aren't built around an enterprise-grade security model from the ground up. In order to both, protect customer data and meet regulatory compliance requirements, insurance organizations need strong database security that does not get in the way of legitimate and authorized access to records.

## Powering a business ontology with semantic data

Simply having the ability to store, access, and protect every type of data isn't enough to make that data actionable. While it's great to have semantic relationships tying records together, these relationships won't help an insurance organization without methods to process and analyze that data.

A business ontology is vocabulary that specifies terms, definitions, and synonyms. Some industries have developed industry-wide standard ontologies, such as Financial Industry Business Ontology (FIBO). A defined business ontology informs the Extract-Transform-Load (ETL) process when importing and transmitting records. However, even inside a single organization, the business ontology can have many variations. This creates difficult obstacles for ETL in traditional relational databases.

A comprehensive insurance data system needs to ensure that data reflects and strengthens the ontology of the business. A strong ontology makes information easily available when needed, whether or not individuals accessing the system are aware of its accessibility or existence. This system – the “enterprise brain” – intelligently discovers connections between semantic relationships to form a knowledge graph. This would provide a number of powerful features, such as:



- Enterprise-wide search with ontology-aware semantic connections
- Automatic geographic information integration
- Intelligent ingestion and linking of ontologies
- Integration of third party data sources to enhance internal records
- Population of machine learning models
- Automatic discovery of new facts based on semantic data

Ontology-aware semantic connections drive new product development, better customer interactions, and deeper knowledge of insurance operations. Better yet, intelligent interaction between semantic relationships and business ontology is key to successfully achieving a 360-degree view of a customer.

### Semantic model case studies

With an enterprise brain backed by MarkLogic's multi-model semantic database system and Wipro's Semantic Data Hub, insurance organizations have seen great improvements in data management.

#### Importing and processing from legacy data sources

Imagine having 70 years' worth of paper documentation and other records in SharePoint, emails, file systems, and a case management solution. This is a common situation for many insurers. If regulators ask you to delete any records that contain Personally Identifiable Information (PII), how do you find records containing PII and properly process them? If you delete the records, do you also delete the knowledge and history they contain?

One insurer processed their paper records by scanning them and applying OCR. These results were ingested into MarkLogic, where PII was successfully identified and removed while

maintaining context and knowledge. Industry ontologies were also applied to map the content themes and codes. Users can now perform semantic searches across all their content, gaining new insights using the historical data.

#### Modernizing and digitizing policies and practices

A large insurance company was interested in modernizing and digitizing its insurance policy processes and customer experience. The company's CSV-formatted customer information, XML-formatted billing records, JSON-formatted policy records, and PDF-formatted claim records were stored across multiple source systems, including relational and mainframe systems. Delivering a complete record of a customer's policy and billing information relied on massive ETL processes to match and merge customer data across multiple sources. A multi-million dollar master data management (MDM) platform was available but it required as much as 18 months of lead time to implement changes to its schema, making it incredibly difficult to build internal and external applications with a comprehensive view of their data.

A Data Hub was implemented to ingest data from all of the different sources. This data was intelligently enriched and harmonized and records were linked based on user-configurable rules.

#### Navigating data sources with an enterprise brain

One obstacle facing insurance is the constant need to manage data from multiple sources in different formats. Researchers in the pharmaceutical and scientific industries face a similar challenge, often needing to navigate to multiple internal data sources. Due to the complex ontologies they utilize, traditional approaches based on data catalogs have not been satisfactory. In addition, the traditional

approaches do not account for the complex nature of data using both structured and unstructured formats and coming from multiple sources. There are multiple terms for individual drugs or chemicals, and data is messy or inconsistent across different repositories. Researchers are often unable to find the right information, or they find conflicting information. The time and effort spent on these systems affects productivity.

By building a semantic metadata management layer, it becomes easier to discover linked documents involving the same items under different names, such as chemicals, diseases, and symptoms. For the insurance industry, this could help link different descriptions of incidents, property, and people.

### Measuring success with a semantic data hub

Undergoing the migration to a new method of storing and managing data is a daunting task for even the most agile insurance organization. Before tackling such a project, it's important to know how an organization will be able to measure the success of their semantic data hub and enterprise brain.

### About MarkLogic

Data integration is one of the most complex IT challenges, and our mission is to simplify it. The MarkLogic Data Hub is a highly differentiated data platform for simplifying data integration, enabling organizations to achieve a 360° view faster than ever. By removing friction at every step of the process, MarkLogic helps organizations gain agility, lower IT costs, and safely share their data.

Organizations around the world trust MarkLogic to handle their mission-critical data, including 6 of the top 10 banks, 5 of the top 10 pharmaceutical companies, 6 of the top 10 publishers, 9 of the 15 major U.S. government agencies, and many more. Headquartered in Silicon Valley, MarkLogic has offices throughout the U.S., Europe, Asia, and Australia. For more information, visit [www.marklogic.com](http://www.marklogic.com).

Explore online to find answers on how the MarkLogic technology can help your organization. <https://www.marklogic.com/solutions/insurance/>

Contact us at [sales@marklogic.com](mailto:sales@marklogic.com)

An insurance organization should see a more accurate and transparent view of customer data over the full engagement lifecycle for sales, marketing, underwriting, risk management, and operations teams.

### Targeting results with a semantic data hub

Many of the following goals make excellent targets for a semantic data hub migration.

- Lower application development costs
- Gain new insights faster on historic data
- Improve performance on database transactions
- Accelerate application delivery
- Discover and make use of proactive customer engagement opportunities
- Automate claims processing
- Improve risk calculation with full view of underwriting data
- Increase product distribution
- Speed up reporting for regulatory compliance
- Leverage cloud capabilities and deployment

● **Wipro Limited**

Doddakannelli, Sarjapur Road,  
Bangalore-560 035,  
India

Tel: +91 (80) 2844 0011

Fax: +91 (80) 2844 0256

**wipro.com**

Wipro Limited (NYSE: WIT, BSE: 507685, NSE: WIPRO) is a leading global information technology, consulting and business process services company. We harness the power of cognitive computing, hyper-automation, robotics, cloud, analytics and emerging technologies to help our clients adapt to the digital world and make them successful. A company recognized globally for its comprehensive portfolio of services, strong commitment to sustainability and good corporate citizenship, we have over 175,000 dedicated employees serving clients across six continents. Together, we discover ideas and connect the dots to build a better and a bold new future.

For more information,  
please write to us at  
**info@wipro.com**

