HOW CAN SEMANTIC INFORMATION MANAGEMENT HELP YOU TO PRESERVE MEANING IN A DYNAMIC DATA ENVIRONMENT?
EXECUTIVE SUMMARY

With information streaming in from more varied sources and at a faster pace than ever before, organizations are having an increasingly difficult time deriving accurate meaning from their data. Data governance systems that were once able to organize and process enterprise information are now too slow and siloed to handle what has become a rapidly evolving and heterogeneous data landscape. As a result, companies may mischaracterize otherwise accurate data, putting their reliability and competitiveness at stake. By connecting all kinds of data and its metadata in a more accessible way than ever, semantic data management systems empower users, data stewards and analysts to make more accurate use of, gain precise and timely insights from, and unlock the true meaning behind their business’s data.

TOP 5 CHALLENGES IN THE EVOLVING DATA LANDSCAPE

A competitive advantage once reserved for massive organizations handling unprecedented amounts of information, businesses of all sizes and stages of growth are now leveraging data to improve their effectiveness. As they aggregate more and more data, however, it quickly becomes unwieldy, leading companies to develop strategies to ensure their data remains accessible, usable and accurate. These efforts, collectively known as data governance, include everything from creating new organizational structures to designing and implementing new processes, policies and technologies to enable them. As the data landscape has evolved, so have the challenges, including:

HOW CAN SEMANTIC INFORMATION MANAGEMENT HELP YOU TO PRESERVE MEANING IN A DYNAMIC DATA ENVIRONMENT?
1. NEW ROLES AND MORE PERSONNEL

Over the last several years, enterprises have invested heavily in their data governance programs. For example, in a 2013 report Gartner estimated that just 50 companies worldwide had created and filled the position of Chief Data Officer. Just over two years later, a Forrester survey of more than 3,000 global data and analytics decision-makers found that 45 percent of their companies had appointed a CDO, representing a massive increase in businesses prioritizing the addition of new personnel along with processes and organizational changes to support their data-driven endeavors.

2. NEW INFRASTRUCTURE

As organizations expand to accommodate the growing streams of data, so do the systems they are using to store it. From data lakes and XML databases to the traditional relational databases and the data dictionaries developed to organize them, companies are attempting to quickly build the infrastructure to support their new data strategies. However, the data landscape has changed significantly over time and continues to evolve rapidly. Success in this new environment requires the flexibility to integrate existing systems and technologies with new ones, as well as more seamless processes and practices for accessing and managing data.

3. NEW SOURCES

Today, data comes in from an unprecedented number of sources at dizzying rates. For example, big data technology is enabling businesses to aggregate information from a growing number of sources - from both outside and within a company. Unequipped for the fast-paced atmosphere that this influx of data has made possible, legacy data management systems often appear sluggish in comparison - their inability to accommodate governing of new data sources and data artifacts renders them ineffective at a time when data environments are increasingly heterogeneous. Meanwhile, data as a whole remains inadequately understood, appreciated and managed in the enterprise. While in theory, being data-rich should make it easier for an organization to undertake new initiatives and efficiently support current operations, in practice, poor understanding of data may often have the opposite effect.
4. INCREASING VARIETY

Data from disparate sources is inherently varied, especially in the ways that it is represented, structured and stored. Even a field as basic as “Name” could differ between two datasets: one could list first and last names together as one item, while another might separate them. As trivial as this difference might seem, it can have major consequences: What might happen to an airline company looking to combine its frequent flyer records after a merger if it cannot properly map the two programs’ data to one another?

With variations like this, typically organizations relying on legacy data governance systems struggle to derive consistent meaning from their data. In fact, more than half of companies report that variety – not volume or velocity – represents the biggest challenge when it comes to managing their data, according to a New Vantage Partners survey.

5. CONSISTENT MEANING

Data is especially valuable to companies when they can share it between departments. However, those that are unprepared to handle variety in their data often find that important meaning is lost, or worse, misaligned when transferring information across work groups.

For example, without a consistent glossary, terms as simple as “customer” may be misconstrued. While it is not necessary for organizations to settle on a single definition of every term, it is crucial that each group working with the data understands how it is used in its various contexts. Without a system in place that can organize these definitions and make them accessible to everyone, companies are forced to rely on the individuals who work with the data day to day to preserve its meaning. While this may be enough for a single department to survive, such a practice is impossible to scale.
CONSEQUENCES OF POORLY MANAGED DATA: “DATA DEBT”

When these challenges are not fully appreciated or addressed, highly-regulated industries whose compliance depends on accurate reporting, or those whose dynamic markets demand agility, find themselves unable to unlock the true potential of their data. In fact, some find themselves bogged down by an overwhelming and ever-growing swamp of data.

In the world of software development, tools accrue “technical debt” as their engineering teams repeatedly choose the fastest, easiest path to resolving bugs. While the resulting software may appear functional, they have added levels of complexity and ignored design flaws beneath the surface, creating a “technical debt” that makes future changes to the program far more costly and time-consuming than they should be. The same can be said for organizations that opt for instant, easy solutions for problems with their data. While they may allow users to perform their jobs in the short term, such expedient practices allow quality and consistency to corrode in the background. Eventually, companies are held accountable for this “data debt,” whether by acknowledging that the quality of their data has deteriorated over time, or worse, by facing catastrophic loss after inaccurate data allows a major transaction to be mischaracterized.
SOLVING THE PROBLEM WITH A SEMANTIC APPROACH

Semantic information management makes it easier for companies to reconcile data from different sources by compiling and organizing information about that data, its metadata. Semantic systems can support this contextual data with the same standards-based representation used to describe the data itself. In this standardized environment, data stewards can create new connections and relationships between data and the additional information needed to describe it, document its uses, sources, and so on. For the first time, data stewards are given direct access to the true meaning behind their data.

As a result, organizations no longer have to rely on a few individuals who know how each department defines a term. Instead, data stewards are able to access, organize and adjust this defining information directly in a single, organized system. Just as they can query the number of customers in a certain department’s records, they are also able to search for and view that dataset’s definition of “customer,” ensuring the inferences they draw from the data are accurate. Additionally, semantic technology allows businesses to make their data and metadata more powerful by providing a set of standards that can connect all information in new ways. Semantic systems can be “instructed” through explicit connections to understand that fields across different departments, such as “name,” “customer name,” and “First Last,” all mean the same thing. That way, data sharing or aggregation can be done in a way truly consistent and relevant.

In this new environment, all people who work with an organization’s data, whether they are data users or data stewards, are far more empowered, enabling faster and more accurate decision-making at every level. Rather than spending days or even weeks waiting for requests or changes to be
routed through busy IT departments, data stewards in a semantic environment can enable governance of new artifacts and capture new metadata on their own. This also frees up valuable time for IT departments for more pressing concerns than adding fields and relationships for the data governance tool.

Organizations whose data consistency and integration are a critical foundation for their regulatory compliance stand to benefit even more. With their metadata maintained and connected within semantically integrated systems, they can automate distribution of their governance datasets and metadata. Moreover, activity trails in semantic systems are far more transparent than in more disparate solutions, making it easier to mitigate risks and substantiate compliance.

To quote one data manager in a multi-national Pharma company describing the benefits this approach brought to his organization:

“We can now finally really link together:

- Data Structures to Data Structures (via Models).
- Data Structures to Governance (with administration and validation).
- Data Structures to Web Services (with configurable schemas).
- Data Structures to Exports (in Excel, PDF and other formats).
- Data Structures to User Interfaces and Reports.
SO WHAT ARE YOUR NEXT STEPS?

Your concerns and priorities may be informed by common questions such as:

1. What is your focus for better data management – is it people, process or technology? Or all of these, but with a need to have them work together better?
2. Is your current way of governing data assets working for you?
3. Are you able to manage information about your data as well as the data itself?
4. Does the lack of consistent semantics in your data lead to errors or bad business decisions?
5. Do you need to integrate new data sources quickly, or adapt frequently to changing business or technology requirements?
6. Does your organization need to better integrate management of business concepts, data and metadata?

SUMMARY

Semantic data management demystifies metadata. By putting all data and metadata on the same playing field, a semantically integrated data strategy gives stakeholders the ability to quickly gather meaning out of their ever-evolving datasets. As a result, their companies become more agile and intelligent, both in their ability to respond and predict. Semantic data management makes data more consistent, reliable, sharable and usable for analysis – an essential goal of any enterprise data governance program.
HOW CAN SEMANTIC INFORMATION MANAGEMENT HELP YOU TO PRESERVE MEANING IN A DYNAMIC DATA ENVIRONMENT?

SOURCES

http://www2.hawaii.edu/~lipyeow/pub/icde07-ontology.pdf
http://www.topquadrant.com/products/topbraid-edg/
https://www.forrester.com/report/Top+Performers+Appoint+Chief+Data+Officers/-/E-RES123064
http://www.gartner.com/newsroom/id/2506315
http://www.slideshare.net/bobdc/semantic-web-standards-and-the-variety-v-of-big-data
ABOUT TOPQUADRANT

As an Enterprise Information Management (EIM) innovator, TopQuadrant helps organizations succeed in data governance. We provide agile solutions for managing information regardless of its structure, origin or location.

Our products use standards-based graph technologies—because connections are important. Making information easy to connect enables unparalleled flexibility for organizing, governing and using it in today's dynamic data environments. TopBraid Enterprise Data Governance™ (EDG) supports the governance and provisioning of enterprise metadata, business glossaries and reference data. TopBraid Reference Data Manager™ is a lighter configuration of TopBraid EDG focused on governance and curation of reference datasets with comprehensive metadata. TopBraid Enterprise Vocabulary Net™ supports collaborative management of taxonomies and ontologies and offers auto-classification of documents and search enhancement. TopQuadrant customers include many government agencies and Fortune 1000 companies in numerous industries including pharmaceutical, financial services, energy and digital media. For more information, visit [www.topquadrant.com](http://www.topquadrant.com).