

# Semantic Alignment

*Why it matters*

Collibra Vision paper

# Abstract

---

In today's business ecosystems, information has become a competitive and strategic asset. Being able to exchange data and to interpret the information in the data that has been exchanged in the right context and within a reasonable time is a top priority for many organizations. Starting from three simple but serious questions regarding data semantics, data utilization, and data governance that pop up daily in information-intensive enterprises, we easily identify a value proposition for semantic alignment. However, current techniques that claim to create semantic alignment in this sense are unsatisfactory, both theoretically and as far as the quality of the results is concerned. They systemically ignore the subtle gap that looms between information sharing among people (i.e. knowledge sharing) at the business/social level on the one hand; and information sharing between computer systems (i.e. data exchange) at the operational/technical level on the other hand. A solution requires organizations to look beyond mere technical fits and think in terms of mechanisms that transcend their IT infrastructure to a sustainable information-centric infrastructure that meaningfully aligns business with IT. To achieve this goal, we pinpoint two essential requirements: business semantics management and data services.

---

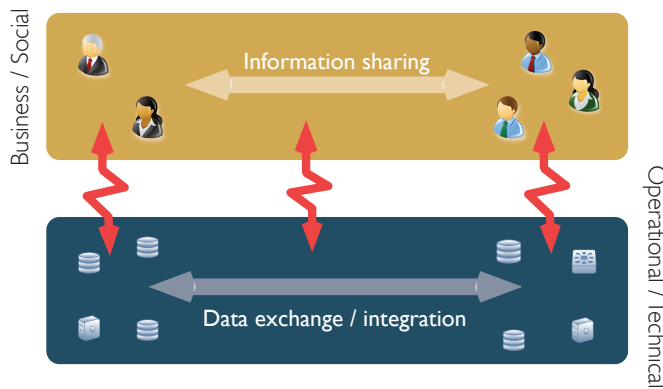
## Table of Contents

<b>Three Simple Questions</b>	<b>1</b>
<b>Semantic Alignment</b>	<b>2</b>
What	2
How	3
Why	6
<b>Conclusions</b>	<b>8</b>
<b>About Collibra</b>	<b>9</b>

---

# Three Simple Questions

In today's business ecosystems, information has become a competitive and strategic asset. Being able to exchange data and to interpret the information in the data that has been exchanged in the right context and within a reasonable time is a top priority for many organizations. This is something that humans manage fluently in their daily face-to-face discourses. However, current techniques are unsatisfying, both theoretically and as far as the quality of the results is concerned. Semantic alignment is hampered by the usually ignored gap that looms between information sharing among people (i.e. knowledge sharing) at the business/social level on the one hand and information sharing between computer systems (i.e. data exchange) at the operational/technical level on the other hand.



This information sharing gap makes it difficult to answer the following three simple but critical questions about any data asset that describes your organization:

## What does my data mean ?

Data has no informational value without meaningful interpretation. It is impossible to share data with business partners or customers. It is difficult to exploit data in order to answer strategical questions,. It is also impossible to compare your data collections and derive intelligence.

*Understanding* what data means is a time-consuming exercise. As the meaning (that is the semantics) is not made explicit this has to be repeated over and over. This is aggravated by the fact that in the course of this "chain of inquiries for data clarification" many colleagues are wrongly interrupted before that single person who actually does understand the data has been spot by chance.

## Where and how is my data used ?

For many organizations, it is far from clear where their data is located and how many copies are available. Moreover they lack any overview of how their data is related to each other, which applications access or manipulate their data. *Maintaining* applications is a time-frittering job. Moreover, most organizations carry the legacy of an information infrastructure that has grown organically over the years, or has gone through a series of refactorings due to organizational restructuring, splits and merges. On top of this, information is usually stored in the minds of employees. While this kind of situation already results in a very inefficient way of handling data, this poses even greater risks when these persons would leave the organization.

*"The market lacks tools capable of rationalizing business process models, logical information models and repository management tools for automated semantic resolution in SOA."*

Gartner, The Emerging Vision for Data Services: Logical and Semantic Management

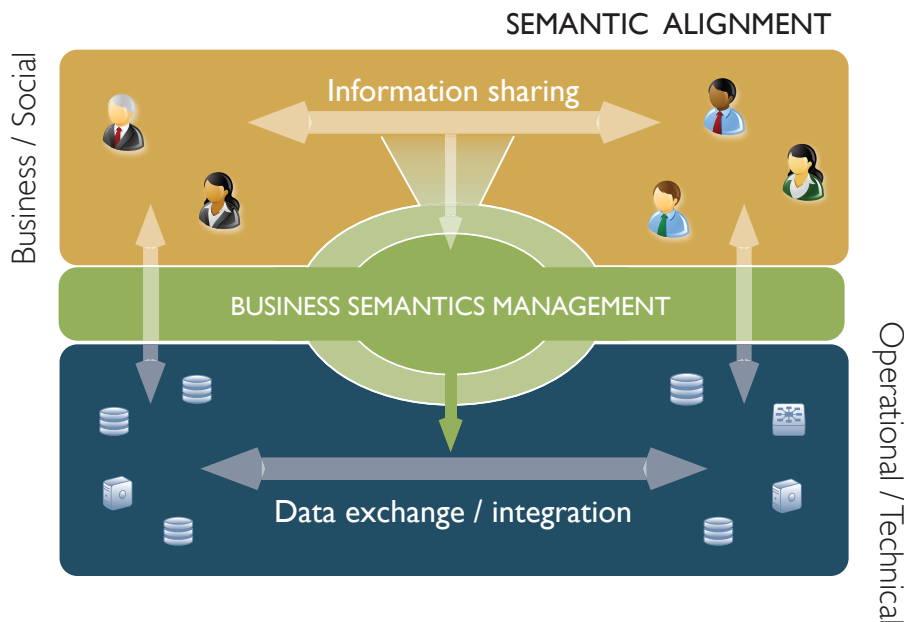
## Who is responsible for my data ?

In today's business ecosystems, trust is considered extremely valuable. Gaining the confidence from your business partners and other stakeholders is achieved uniquely through data governance. Being able to pinpoint a person that holds ownership for a certain data asset, knowing how this asset has evolved over time, knowing and being able to track who has changed what and when, are all abilities that are key for genuine governance.

# Semantic Alignment

## What is semantic alignment ?

Collibra is convinced that bridging the gap between the business/social layer and the technical/operational layer of an organization would reduce daily operational costs and create new opportunities for value creation, both now and in the future. Collibra coins *semantic alignment* as "the ability to efficiently exchange data between the technical/operational layer and the business/social layer and to correctly interpret the meaning of the data that has been exchanged in the right context and within a reasonable time".



As illustrated above, semantic alignment creates added value at three different dimensions

1. It empowers *knowledge sharing* between business stakeholders through more accurate information delivery. For example, semantic alignment enables better search, navigation, discovery, content management, web-sites, and many other knowledge-intensive applications.
2. It empowers *data exchange* between disparate systems as it takes care of the automatic transformation between data formats. This provides a better foundation for many other initiatives such as master data management, business intelligence, SOA, BPM, etc.
3. It enables the *alignment of business and IT* by explicating the meaning, usage and whereabouts of all data assets.

## How can we achieve semantic alignment

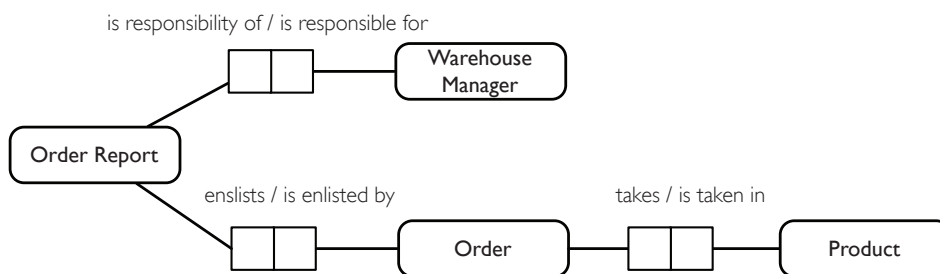
Collibra coins *business semantics management* as an approach to bring business partners together to realize the reconciliation of their heterogeneous metadata; and consequently the application of the derived business semantics patterns to establish semantic alignment.

Business semantics management does for information what business process management does for processes. Semantic Alignment indeed has many similarities to what business process management aims to do for business processes. Another, more technical term for business semantics management would be business metadata management.

### Business Semantics

Business semantics are (business) metadata that describe the information concepts that live within the organization. An important difference with other approaches is that Collibra's business semantics are modeled according to a fact-oriented paradigm that was introduced by the conceptual modeling approach NIAM, the predecessor of Object Role Modeling (ORM). The use of natural language, example populations, and the description of information in terms of elementary facts enhances the potential for re-use and design scalability during business semantics management. The following figure shows an arbitrary semantic pattern composed of three elementary fact types that illustrates the graphical power and simplicity of Collibra's fact-oriented approach.

The following figure gives an example of the power and simplicity of the fact-oriented approach.



The above representation is extremely easy to read. From left to right, the above pattern reads:

- An **Order Report** *is the responsibility of* a **Warehouse Manager**
- An **Order Report** *enlists* an **Order** which *takes* a **Product**.

---

*"I am impressed. The difference between Collibra's solution and what I have seen and used in the past is crystal clear to me, and it really does work.*

*You can take a concept and make the contextual meaning clear, it provides insight and understanding, it reduced complexity, ... It is better !"*

Patrice Krakow, System Integration Coordinator, SCA Packaging

---

*"With the emergence of information as a valuable asset, it is metadata that becomes key to the valuation and leveraging that value, yet most organizations manage meta-data in a very ad hoc manner."*

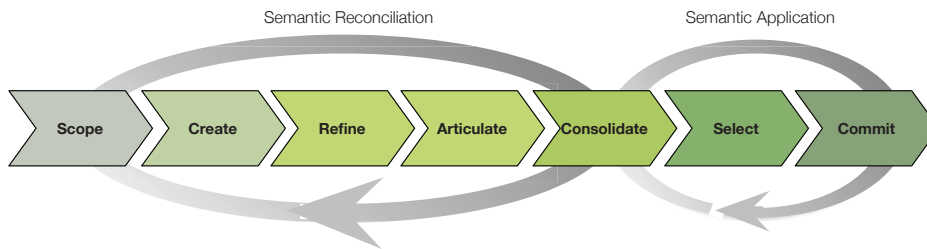
Mark A. Beyer, Gartner

*Business semantics provide a kind of abstracted, federated, and virtualized way to access and deliver data in a more efficient and aligned manner.*

## Methodology

By default, business semantics serve “open” information systems, and hence the requirements and limitations of semantic alignment cannot be entirely known before completion. In contrast to waterfall-like approaches that focus on a broad design upfront, agile methods perform short milestone driven revision iterations in order to cope with dynamic environments such as the extended enterprise. Collibra’s full-cycle business semantics management is established by two operational cycles each grouping a number of activities. This is illustrated below.

1. **Semantic Reconciliation** is the first cycle of the full-cycle business semantics management methodology. In this phase, business semantics are modeled by extracting, refining, articulating and consolidating fact-types from existing sources such as natural language descriptions, existing metadata, etc. Ultimately, this results in a number of consolidated language-neutral semantic patterns that are articulated with informal meaning descriptions (e.g., WordNet word senses). These patterns are reusable for constructing various semantic applications.



2. **Semantic Application** is the second cycle of Collibra’s full-cycle business semantics management. During this cycle, existing information sources and services are committed to a selection of semantic patterns. This is done by selecting the relevant patterns, constraining their interpretation and finally mapping (or committing) the selection on the existing data sources. In other words, a commitment creates a bidirectional link between the existing data sources and services and the business semantics that describe the information assets of an organization. The existing data itself is not moved. On the contrary, the business semantics provide a kind of abstraction layer to access and deliver this data in a more efficient and aligned manner.

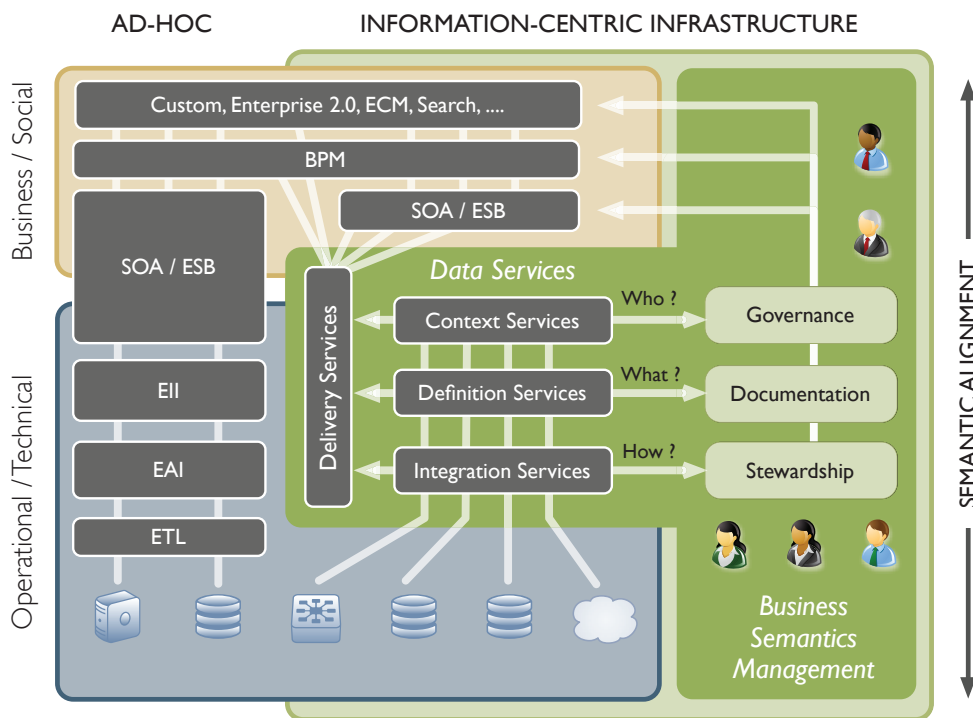
## Architecture & Positioning

How does Collibra’s semantic alignment vision position itself within the existing IT landscape? Collibra’s semantic alignment vision is similar to what Gartner labels as the “Information-centric Infrastructure”. Collibra is convinced that there are two technologies necessary to implement this vision: business semantics management and data services. The image below provides an overall conceptual perspective on the move from an ad-hoc infrastructure to an information-centric infrastructure to achieve semantic alignment.

*An information-centric infrastructure is the alignment of metadata, standards, content formats and applications to support consistent and seamless enterprise-wide information capture, persistence, transformation and delivery.*

Gartner





## Business Semantics Management

Business semantics management is the human-driven part of implementing the semantic alignment vision. Using Collibra's business semantics, existing disparate data and service sources are annotated with rich semantic patterns that establish the meaning of the information assets. It also supports other human-centric involvement such as governance and stewardship. Business semantics management empowers all stakeholders in the organization by a consistent and aligned definition of the important information assets of the organization. The available business semantics can be leveraged in the so-called business/social layer of the organization. For example they can be combined with a content management application to provide a consistent business vocabulary and enable better navigation or archiving of documentation. This can be further complemented by enterprise search engines, and richer Semantic-Web ready websites, etc.

Service-oriented integration (SOI) is defined as integrating computing entities using only service interactions in a service-oriented architecture. Service-oriented integration addresses problems with integrating legacy and inflexible heterogeneous systems by enabling IT organizations to offer the functionality locked in existing applications as reusable services.

In contrast to traditional enterprise application integration (EAI), the significant characteristics of services-oriented integration are:

**Well-defined, standardized interfaces** – Consumers are provided with easily-understood and consistent access to the underlying service.

**Opacity** – The technology and location of the application providing the functionality is hidden behind the service interface. In fact, there is no need for a fixed services provider.

**Flexibility** – Both the providers of services and consumers of services can change - the service description is the only constant. Provided both the provider and consumer continue to adhere to the service description, the applications will continue to work.

## Data Services

Data services replace the old practices such as EAI (Enterprise Application Integration), EII (Enterprise Information Integration), B2B Integration, and ETL (Extract Transform Load). Technically, they fit better in a Service Oriented Architecture and is hence often called service-oriented integration. Collibra's Information Enabler is a semantically empowered data service providing more than mere integration services.

Applications can leverage the business semantics to enable semantically rich data services including context services (how, where and by whom is this information asset used), definition services (what does this information concept mean in this particular context) and integration services (where is this information stored, transform a piece of data from one format to another, ...). This enables these data services to deliver the right information at the right time in the right context and in the required format.

## Why would you need semantic alignment ?

The benefits of semantic alignment are twofold:

### Documentation

Semantic alignment makes it possible for any stakeholder to answer the following three crucial questions about their data: (i) what does it mean; (ii) where and how is it utilized; and (iii) who is responsible for it?

How does such a comprehensive overview on your enterprise information creates value for your business:

- **Reduced operating costs:** with a complete information overview, the chain of inquiries for data clarification is maximally reduced. Hence, this minimizes operations, cuts cost and time, and lowers the frustration caused by repeated data misinterpretations.
- **Reduced maintenance costs:** using Collibra's business semantics management approach, the existing (legacy) data sources and services are well-documented in a way that every stakeholder understands. Furthermore, because this bidirectional link is also used operationally in the actual information systems, this form of documentation always remains up to date.
- **Faster time to market:** because your organization has a better understanding of the meaning, the whereabouts, and the utilization of its data assets, it can adapt and extend its systems seamlessly to address new market requirements.
- **Faster integration with M&As:** one of the major hurdles to overcome when integrate two large and heterogeneous environments is to know what data assets exists, how they are used and stored, and who is responsible for them. Having a comprehensive overview of this information makes a smoother merging of departments or organizations possible.
- **Better information governance:** knowing how and where your information assets are utilized and stored. Knowing who is responsible. Knowing how they have evolved over time, who has changed what. Knowing this turns organizations towards true information governance.

*A recent survey indicates that strategic application spending in the area of Data Governance is expected to triple over the next two years.*

DAMA

(<http://www.irmuk.co.uk/dg2009/>)



---

***“Despite the great promise of SOA, challenges remain. Semantic integration – or the integration of differing types and structures of data to create meaning – will emerge as a significant problem.”***

Accenture, Major trends that will shape IT

([http://www.digitalforum.accenture.com/DigitalForum/Global/ViewByTopic/Economy/o6o7\\_major\\_trends\\_shape\\_it.htm](http://www.digitalforum.accenture.com/DigitalForum/Global/ViewByTopic/Economy/o6o7_major_trends_shape_it.htm))

---

## **Operational and Technical Efficiency**

Knowing the answer to the above three questions would not only be valuable for human stakeholders. Collibra has developed a product suite that can fully operationalize the business semantic models by committing them to the underlying (legacy) data sources or services. This provides the following benefits:

- **Automatic data transformation:** Collibra has created a run-time server, called Collibra Information Enabler, that automatically transforms data from one format to another as long as these formats are documented by business semantics. This is the first step towards a linearly scaling solution for an exponentially scaling problem.
- **Better SOA / BPM:** Through Service Oriented Architecture and Business Process Management, many organizations have tried to introduce conceptual insight, understanding, and flexibility to their enterprise architecture. However, information still remains the foundation on which these initiatives are built. Many experts agree that a lot of investments into SOA and BPM will not provide the promised benefits when people forget about the data semantics. Collibra’s business semantics solutions provides the unique straight way to support these initiatives and actually deliver on its promises.
- **Enterprise 2.0:** The recent trend towards Enterprise 2.0 gives power to the users to better share information with their peers and use enterprise applications like they use their consumer applications at home. For the IT department, it is not straightforward to support this trend and embed it safely into the overall enterprise architecture. Through semantic alignment, however, business users are empowered to create custom applications (also called mashups) to efficiently share and deliver information. The information assets from the organization are made available for them through business semantics in a way they understand so they can use it without the involvement of IT. However, IT is still in charge to map (or commit) these business semantics to the underlying source data and services.
- **Cloud Computing:** Another emerging trends is cloud computing. It is however highly undesirable to directly connect data that is stored in the cloud to the processes and business logic of the organization. Semantic alignment, creating an abstraction layer between both, is the ideal solution to use and leverage information from the cloud safely and efficiently into the organization.
- **An information-centric infrastructure:** Gartner describes an information-centric architecture as the alignment of metadata, standards, content formats, and applications to support consistent and seamless enterprise-wide information capture, persistence, transformation and delivery. Gartner envisions the information-centric architecture as a key building block for Enterprise Information Management. Collibra’s Business Semantics Management approach is the only way for organizations to convert smoothly to such an information-centric infrastructure.
- **Outsourcing:** Maximize the ability to outsource by keeping the business knowledge local through information models, policies, rules and decouple these from implementation code.

# Conclusions

---

Being able to exchange data and to interpret the information in the data that has been exchanged in the right context and within a reasonable time is a top priority for many organizations. In this paper, we set a value proposition for semantic alignment. In terms of three simple but serious questions about data semantics, utilization, and governance typically posed in information-intensive organizations.

Current techniques that claim to create semantic alignment in this sense are disappointing, both theoretically and as far as the quality of the results is concerned. They deny the subtle gap that looms between information sharing among people (i.e. knowledge sharing) at the business/social level on the one hand; and information sharing between computer systems (i.e. data exchange) at the operational/technical level on the other hand.

A solution requires organizations to look beyond mere technical fits and think in terms of mechanisms that transcend their IT infrastructure(s) towards a sustainable information-centric infrastructure that aligns business with IT. To this end, semantic alignment must be implemented with methods and tools for business semantics Management and data services. Through accurate data understanding and governance, and knowledge sharing, organizations are empowered to optimize their operations. Moreover, top-line growth is guaranteed by leveraging previous investments, improving development of new applications, and ultimately increasing shareholder value.

# About Collibra

---

Collibra is an independent enterprise software company and spin-off from the Laboratory for Semantic Technology and Application Research (STARLab) at the Vrije Universiteit Brussel. Collibra's mission is to deliver solutions for information governance, sharing and delivery throughout the extended enterprise. Similarly to business process management, Collibra coins its technology business semantics management.

Collibra consolidates more than 10 years of application-oriented semantic technology research at one of the leading centers of excellence in its field. Given this technological advantage, Collibra is convinced it can provide its customers with the best solutions for their information problems.

Collibra has closed a seed round of € 850.000 from VCs. The team expanded with world-class employees and board of directors, including Tony Mary (ex-CEO IBM Belgium, KPMG, Bull, ...) and Dirk Boogmans (ex-CEO Gimv). Collibra has several projects running with large multinationals and government organizations. With its unique technology, backed by a solid methodology and via a strong partner network, Collibra has all the necessary qualities to become market leader in its field.



Collibra Studio is a tool suite that enables IT professionals and business analysts to work together to reconcile and apply the semantics of existing information sources. Business semantics models are created that provide unambiguous definition, identifications and mappings towards the existing data sources.



The Collibra Platform is a kind of repository that enables collaborative Business semantics management. On the Collibra Platform, the information assets within the company can be managed and governed. For a sustainable solution, it is crucial that different stakeholders can collaborate on the meaning and structure of these business semantics.



The Collibra Information Enabler is a run-time server that leverages the business semantics to provide real time data services. The Information Enabler makes automatic semantic integration possible: given a set of business semantics, the Information Enabler can automatically transform data between different formats.

