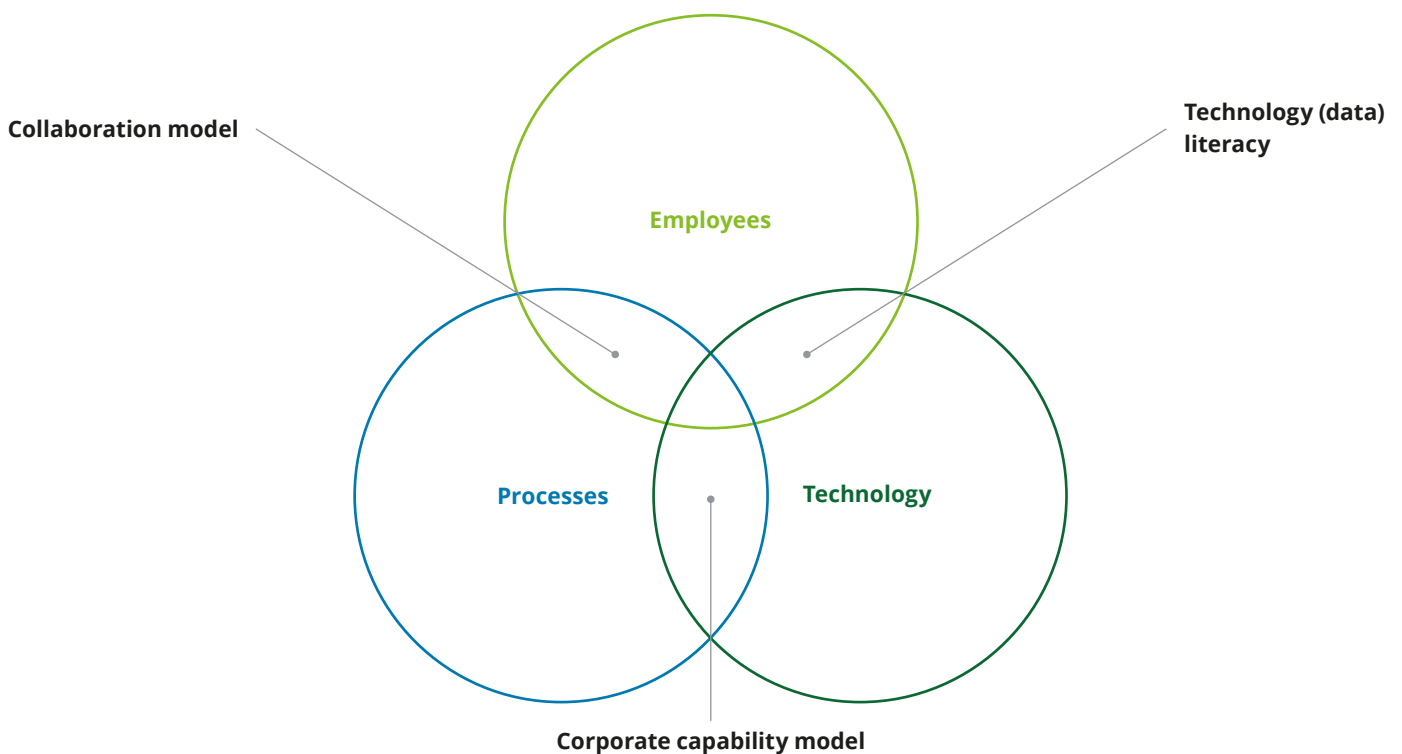# Deloitte.

Using ontologies to
understand and use data
easier and more efficiently

March 2023

Introduction

Modern enterprises leverage data and AI to keep up with the high dynamic of an ever-changing business world. On the way to achieve that goal almost all of our clients face the same challenge: To transform their strategy, operating model and way of doing business such that data is used to drive the corporate capabilities used in processes that generate value.

One of the recent concepts of achieving this is the data mesh, that shifts responsibilities towards the business to produce data products that are standardized in their creation as well as in their access. The data mesh paradigm, however, is not a purely technological invention, but rather a socio-technological endeavor that brings together technology, employees, and corporate processes. The classical Venn diagram of these three is given below. We have added the usually omitted descriptions of the links between the adjacent pairs: Employees possessing an intuition for the capabilities and limitations of technology ("technology literacy"), the enterprise being able to use technology in processes ("capability model") and the employees working together along the process landscape ("collaboration model").

Collaboration model

Technology (data) literacy

Employees

Processes

Technology

Corporate capability model

Achieving literacy, data capabilities and collaboration are at the core of a successful transformation. Here, Deloitte presents a series of papers that explain key aspects how to achieve these three along the transformation towards becoming a data driven enterprise. The series is structured into strategic, tactical and operational aspects of data driven work.

Beginning with the strategy framework we are working along, we introduce our orchestrator for the data transformation journey. As the major tactical pillars of the transformation we focus on the required governance as well as the data-centric process landscape in two further articles.

These concepts are underpinned by operational tools such as data catalogs, data quality and IT platforms which we are also covering in an article. Since these developments need to be sustained by specialized change management, a separate article is dedicated to this topic.

The journey to a data-centric enterprise is a complex transformation that continues to bring new challenges and insights. We will continue to expand and add to our series of articles.

Using ontologies to understand and use data easier and more efficiently

A key feature of a data-driven business is to improve the entire value chain using data and insights generated from data analyses. To do so, a high degree of data quality and a well-founded understanding of the context of data and their business semantics are relevant drivers. Here, context of data indicates, for example, that several columns represent a common topic, e.g., customer address and sociodemographic information. In contrast, semantics describes domain knowledge that is not necessarily stored within the data, e.g., expertise on the economics of currency conversion rates in their interaction with the company's financials. Consequently, to use both data and its semantics in a computerized fashion, domain knowledge connected to the data must be stored in an abstract, machine readable, but humanly usable representation.

A common technique to capture such semantic knowledge is the use of ontologies: formalized descriptions that capture relations between business entities and their abstract realization as data. Adopting ontologies as descriptive tools in data product creation has proven very useful, particularly when connected to a data catalog to provide a holistic understanding of data from all perspectives. As a natural consequence the knowledge available in a company and then stored in an ontology can easily be made available and searchable through knowledge graphs.

More precisely, ontology is a philosophical discipline interested in identifying what exists (ancient Greek ὄντος, 'existing'). An ontology practically describes *what exists* in the corporate environment by four components (see the below figure for the example described in the text):

1. the context of the described entity,

2. the entity's meaning,

3. relations to the (business) environment,

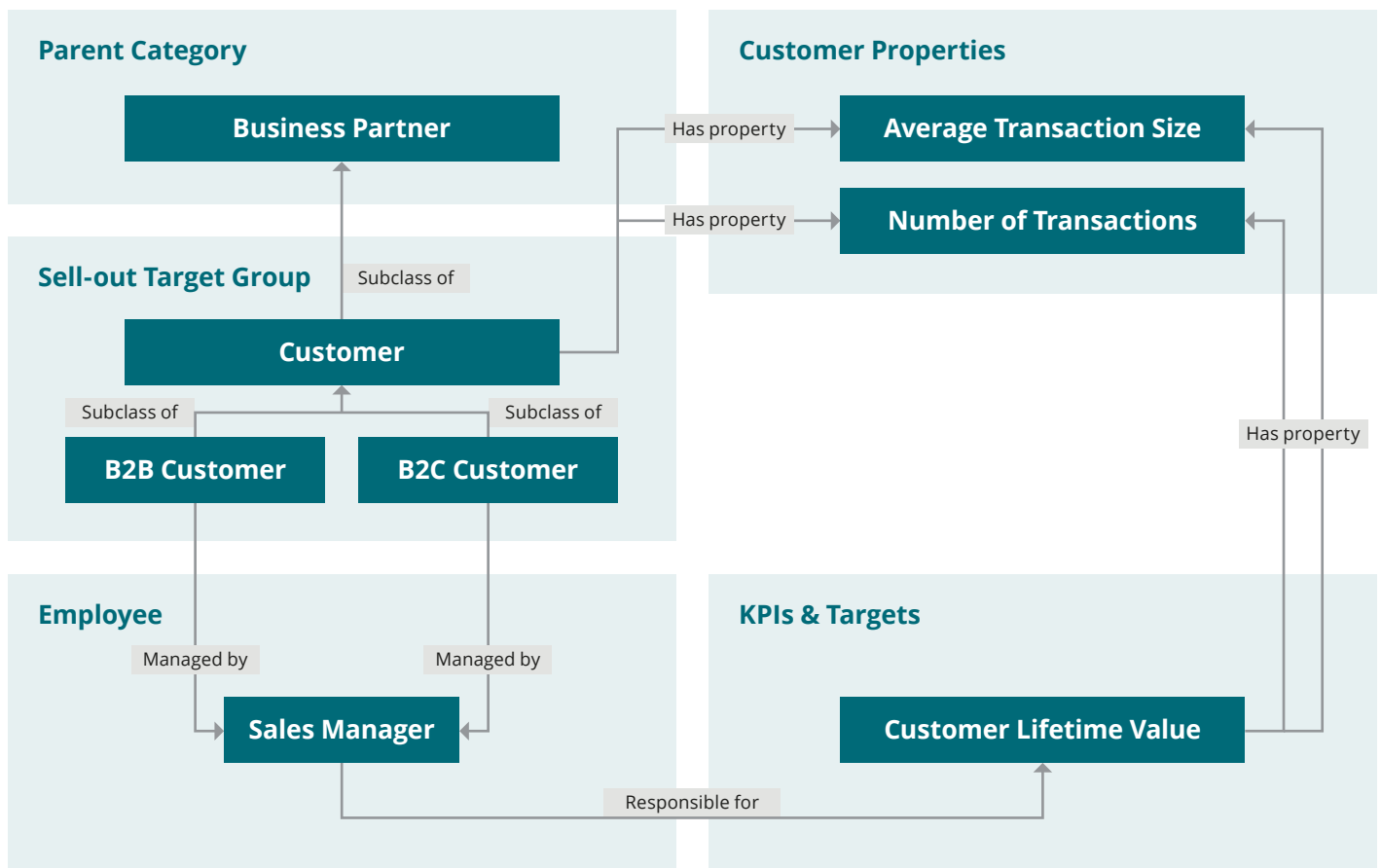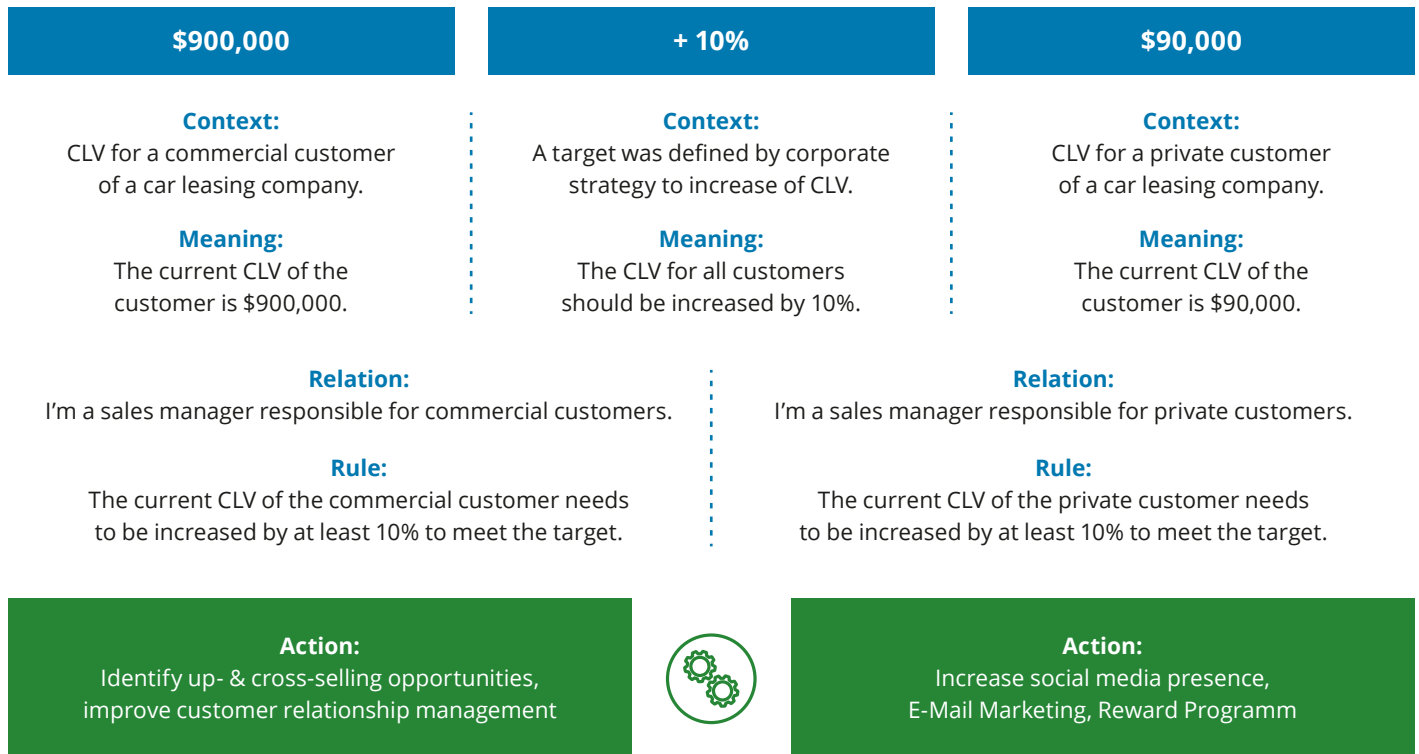4. a business rule to base decisions on.

Typically, these four components are abstractly described in programming as classes, relations between classes, and properties. Classes are abstract templates used to define the structure of objects (*'what exists'* in the corporate environment) that have similar properties. These properties may not yet be known at the time of definition. For example, a class could be 'business partners' (anyone or anything the company collaborates with). A subclass of business partner might be 'customers', since they are certainly important partners, but differ from other subclasses such as 'suppliers' in the way that their contribution to the company is inherently different. Subclasses may have subclasses by themselves such as B2B customers or B2C customers. The subclass can retain some of the characteristics of the parent class (a concept called 'inheritance' in programming) and may have additional properties. In essence, every class or subclass provides a context for its entities.

In practice, using ontologies in a business setting requires a two-stage process: Developing the abstract class system (that may be a living system being subject to further development and change) and instantiating specific instances of these classes as realizations of the classes.

For example, using Customer Lifetime Value (CLV) as a steering mechanism in a car leasing company implies that KPI differ in the contexts of B2B and B2C customers even though the meaning of CLV is identical. CLV for B2B customers is frequently related to leasing an entire fleet and depends on macro-economic drivers and corporate finances, while B2C customers are privately leasing individual cars that fit to their personal stage of their life. Specific customers are established as instances of the B2B or B2C class. Consequently, such context is highly relevant to derive appropriate business rules to maximize the CLV of the customer base.

Creating a large number of interacting classes with connected relations creates significant complexity. We therefore recommend expanding the data governance structure of a corporation to govern not only data, but the four components of ontology to contextualize data using knowledge. To simplify this process, a common method for storage and visualization of all knowledge established in the ontology is employed. This knowledge graph is created immediately after instantiating objects.

Example:
Customer Lifetime Value (CLV)

$900,000	+ 10%	$90,000

Context:
CLV for a commercial customer of a car leasing company.

Context:
A target was defined by corporate strategy to increase of CLV.

Context:
CLV for a private customer of a car leasing company.

Meaning:
The current CLV of the customer is $900,000.

Meaning:
The CLV for all customers should be increased by 10%.

Meaning:
The current CLV of the customer is $90,000.

Relation:
I'm a sales manager responsible for commercial customers.

Relation:
I'm a sales manager responsible for private customers.

Rule:
The current CLV of the commercial customer needs to be increased by at least 10% to meet the target.

Rule:
The current CLV of the private customer needs to be increased by at least 10% to meet the target.

Action:
Identify up- & cross-selling opportunities, improve customer relationship management

Action:
Increase social media presence, E-Mail Marketing, Reward Programm

Parent Category

Business Partner

Customer Properties

Average Transaction Size

Number of Transactions

Has property

Has property

Sell-out Target Group

Subclass of

Customer

Subclass of   Subclass of

B2B Customer   B2C Customer

Employee

Managed by   Managed by

Sales Manager

Has property

KPIs & Targets

Customer Lifetime Value

Responsible for

Such a graph can be applied in multiple ways, four of which we will highlight in the following: Use in data products, integration in a data catalog, to study data quality, and for AI applications.

During the design phase of the data product lifecycle process (see our whitepaper on data products), ontologies are used to capture the conceptual data model and the relationship between data products. One key capability of data products is discoverability. This can only be achieved by embedding an ontology within a data catalog. Additionally, an ontology simplifies the development of new data products since relevant relationships between data and data products are traceable. If there is a culture of data experimentation in the company, the captured ontology itself can also provide new insights and serve as input for the use case process.

Secondly, ontologies improve the efficiency of working with data by making implicit knowledge explicit and storing it in a machine-readable format. When combined with a data catalog, metadata is enriched with the ontology's domain knowledge shaping the single point of truth by connecting entities via their semantic and static information. This fosters data discoverability by providing better understanding of data availability, data storage location and its value-adding contextualization.

Third, ontologies are useful to ensure contextual data quality (see our whitepaper on data quality). The meaningfulness of data for a certain use case depends strongly on the availability and transparency of the data's meaning and context. Additionally, managing data quality combined with meaningfulness of analysis results is assisted by contextual and semantic information. Therefore, clarifying ontological structures and establishing specific instances in the corporate ontology enhances standardization and facilitates topical exchange.

Fourth, the usage of domain knowledge represented by an ontology is particularly useful for machine learning. It enables data scientists to understand the context of a machine learning problem to be solved and, consequently, it allows to easily identify the required data to solve the problem. Using the example given before, a machine learning problem could be to predict the CLV for a car leasing company. The relevant features (variables) heavily differ for customers in the context of B2B and B2C business. Even though the statistical target is to predict the CLV as a number, the business objective is to identify measures to increase sales. Hence, the data scientist is not interested in variables correlating with CLV, but causally impacting it in order to derive actionable insights. For B2B customers macroeconomic variables such as the performance of an industry or certain companies may be considered whereas for B2C customers the individual stages of life such as marriages, childbirth or buying a house are relevant.

Therefore, an ontology allows causality-driven machine learning by enabling a data scientist to separate causal variables and confounders for solving a machine learning problem. Furthermore, it improves the explanatory power of an analysis as it allows for causal interpretation of results, therefore increasing data meaningfulness and quality. Leveraging the previous example, the car leasing company might be using a prediction model to forecast the number of signed leasing contracts until the end of the fiscal year to indicate if sales targets can be fulfilled. While the leasing target numbers of sales agents correlate very well with the number of let vehicles, the target itself is not causally related to the number of vehicles, but the willingness of agents to incorrectly register additionally let vehicles in the coming period, if their current targets are already fulfilled, is.

Finally, many companies currently face challenges caused by demographic changes: In the upcoming years employees will retire more frequently than graduates will join, an effect frequently described as 'war for talent'. As a qualitative consequence, experience and knowledge will be lost. In such a situation, capturing and formalizing expert knowledge about a certain domain is highly beneficial to retain knowledge in a company while replacing retired employees. The use of ontologies and knowledge graphs therefore has significant advantages in their capability to store and retrieve domain knowledge in an understandable format.

Contributors

Mark Arndt
Consultant
Strategy, Analytics and M&A
+49 151 1917 5124
maarndt@deloitte.de

Dr. Max-Hendrik Böttcher
Manager
Cyber & Strategic Risk
+49 151 5448 3572
maboettcher@deloitte.de

Anica Buchholz
Manager
Human Capital
+49 151 5448 4563
abuchholz@deloitte.de

Thiemo Eberle
Senior Consultant
Enterprise Performance
+49 151 1829 5033
teberle@deloitte.de

Frank Eisenhauer
Director
Strategy, Analytics and M&A
+49 151 5448 3868
feisenhauer@deloitte.de

Dr. Christoph Euler
Senior Manager
Strategy, Analytics and M&A
+49 151 1268 3336
ceuler@deloitte.de

Markus Hafner
Senior Consultant
Strategy, Analytics and M&A
+49 151 5448 4069
mhafner@deloitte.de

Kai-Uwe Hess
Partner
Enterprise Performance
+49 151 1829 4406
kahess@deloitte.de

Margaretha Hirsch
Senior Consultant
Strategy, Analytics and M&A
+49 151 5448 4071
mhirsch@deloitte.de

Paul Holbein
Senior Consultant
Strategy, Analytics and M&A
+49 151 5807 6346
pholbein@deloitte.de

Martina Hornung
Partner
Human Capital
+49 151 5807 3900
mhornung@deloitte.de

Dr. Benjamin Klör
Manager
Enterprise Performance
+49 151 5448 4579
bkloer@deloitte.de

Dr. Sebastian Olbrich
Partner
Strategy, Analytics and M&A
+49 151 1488 0375
solbrich@deloitte.de

Ralf Pierson
Senior Consultant
Human Capital
+49 151 2103 3858
rpierson@deloitte.de

Christian Roese
Senior Specialist
Strategy, Analytics and M&A
+49 151 5807 6239
croese@deloitte.de

Marco Ruetten
Manager
Enterprise Performance
+49 151 5807 5861
maruetten@deloitte.de

Jan Scherpinski
Senior Manager
Strategy, Analytics and M&A
+49 151 5448 3411
jscherpinski@deloitte.de

Friedemann Schestag
Consultant
Strategy, Analytics and M&A
+49 151 1268 3791
fschestag@deloitte.de

Tim Staab
Consultant
Strategy, Analytics and M&A
+49 151 1488 1747
tistaab@deloitte.de

Julian Steiner
Senior Consultant
Human Capital
+49 151 5807 4575
julsteiner@deloitte.de

Aizhan Suleimenova
Consultant
Strategy, Analytics and M&A
+49 151 5448 3524
asuleimen@deloitte.de

Tobias Wätzig
Senior Manager
Strategy, Analytics and M&A
+49 151 5800 1565
twaetzig@deloitte.de

Marc Willemsen
Consultant
Strategy, Analytics and M&A
+49 151 1488 0730
mawillemsen@deloitte.de

Jan Wohde
Consultant
Human Capital
+49 151 5448 4643
jwohde@deloitte.de

Contacts

Frank Eisenhauer
Director
Strategy, Analytics and M&A
+49 151 5448 3868
feisenhauer@deloitte.de
Frank is a Director in Deloitte's AI & Data offering with a focus on Data-Strategy, Data-Transformation and Data-Management. He designs and implements data-driven-transformations with an objective on data-value-creation with a holistic and comprehensive approach. In his opinion, data transformation is a team sport.

Dr. Christoph Euler
Senior Manager
Strategy, Analytics and M&A
+49 151 1268 3336
ceuler@deloitte.de
Christoph is a data strategy consultant and data scientist, combining the two in his endeavor to help enterprises transform into data-ready organizations. He has a particular focus on the CEO and CDO organization's operating model design, data / machine learning governance, collaboration models and data literacy. In his opinion, data is a people business - technology is more readily available.

Dr. Benjamin Kloer
Manager
Enterprise Performance
+49 151 5448 4579
bkloer@deloitte.de
Benjamin is working as a Manager for Technology Strategy & Transformation at Deloitte. He acted in different roles to coordinate and deliver transformation projects in various industries for building the data-driven company. Benjamin's professional interests are focused on the full lifecycle of data management, including data strategy definition, data governance design, as well as data value creation by the means of data monetization and modernization.

Glossary

Data Mesh

The data mesh is a domain-driven socio-technological approach for creating decentralized data architectures. It is based on decentral governance structures as a foundation for generating sustainable business value using standardized and re-usable data products. It relies on a flexible collaboration model accross the entire enterprise.

Data Product

A data product is a set of data that is made available for the usage of employees or systems via a standardized API on a marketplace. Its purpose is to realize use cases and therefore to enable the implementation of data-driven services.

Data as a product

Synonymous to Data Product.

Use Case

A use case creates business value by fulfilling an explicit objective. Use cases are based on existing Data Products.

Data Catalog

A data catalog is the central inventory for all data assets within the company. It is made understandable via a glossary of frequently used terms and by highlighting the technical and business data lineage as well as transformation logic.

Data Governance

Data Governance is the discipline that connects data processes, and corresponding roles and responsibilities by formulating binding enterprise-wide policies.

Ontology

Ontologies are formalized descriptions that capture relations between business entities and their ab-stract realization as data.

Data Domain

A data domain takes ownership of data relevant to a common area of interest and implements roles that are responsible for expanding and maintaining the usability of this data.

Deloitte.