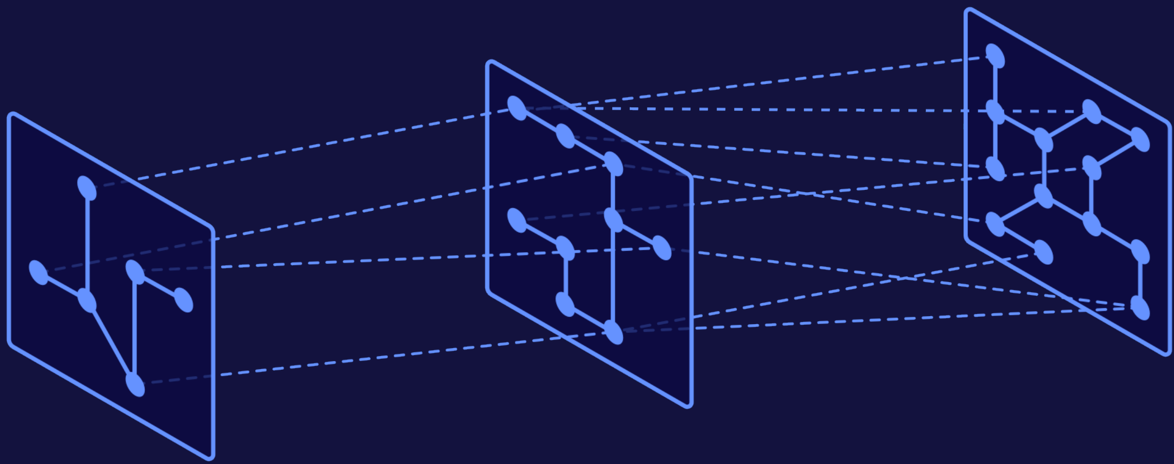
From Master Data to Metadata & Contextual Mastery

By Metagraph Technology



Stefan Dageson

CTO

Introduction

Innovative use of metadata—data about data—provides entirely new opportunities for rich and high-quality information that can both support complex businesses in their daily operations and be instantly utilized in data science, analysis, and AI.

Moreover, metagraphs and dynamic metadata can significantly enhance the application of Large Learning Models (LLMs) and other machine learning techniques through the use of leveraged knowledge graphs and ontologies.

It's all about mastering setup data and managing contextual complexity – at all levels of abstraction within an organization.

This paper elaborates on various types of setup data, including master data, reference data, and metadata, along with their interrelationships. It explores how metadata and ontologies can be dynamically managed and evolved in a scalable manner based on knowledge metagraphs and metagraph technology.

 INORIGO AB | Södra Förstadsgatan 31 | 211 43 Malmö, Sweden | Phone: +46 40 699 500 | www.inorigo.com

Contents

Summary

Navigating the ever-growing complexity of businesses and their value propositions is becoming increasingly difficult. Clearly, much of this difficulty can be attributed to poor data quality. Has too little time been spent on interpreting, cleaning and merging big data? Quite the opposite, there has been insufficient effort in developing small setup data such as master data, reference data and meta data. This setup data is essential as it includes the fundamental identifiers, definitions, and contexts needed for creating and processing all downstream volume data.

To enhance quality, ensure traceability, manage complexity, and utilize AI more effectively, it is essential to undertake substantial groundwork in setting up foundational data. We refer to this setup data as the building blocks that form the basis for identifying and defining important fundamental entities and their contextual interdependencies. These are the foundations upon which larger volumes of data can be built in the value chain.

Everything starts with setting up data, which often receives insufficient attention despite the importance of its quality. To ensure data reflects reality as accurately as possible, we must meticulously manage the creation and governance of setup data at all critical levels. This includes ensuring unique identifiers, eliminating duplicates, maintaining accurate classifications, clearly defining characteristics, keeping categorizations clean, structuring compositions well, and effectively managing cross-dependencies.

Surprisingly often, data is referred to as ‘raw,’ as if it were a natural resource independent of human judgment. Yet, the initial collection and interpretation of real-world analog signals and observations, as well as the process of converting them into digital format, are significantly influenced by human decisions.

These include the specific aspects to assess, when and where to take measurements, and which methodologies and channels to use. Consequently, the initial registration of data inevitably influences the quality of conclusions drawn from all subsequent data sets and aggregations.

INORIGO AB | Södra Förstadsgatan 31 | 211 43 Malmö, Sweden | Phone: +46 40 699 500 | www.inorigo.com

This highlights why not only master data and reference data are becoming increasingly important, but also the broader context in which they are represented. New technology enables dynamic definition of metadata—data about data—and facilitates the connection between master data, reference data, and related event-oriented data to more accurately represent the broader business environment in which it is created.

This new approach to dynamically managing metadata significantly differs from the approach of activated metadata. Activated metadata involves extracting metadata from existing systems and data sources to create a comprehensive and eclectic representation of business semantics based on these sources.

In the activated metadata approach, the setup data and a schema for interpreting semantics and structure are established after the business operational processes have been executed, but before the analysis of results and learning from them. In the emerging dynamic metadata approach, the essential setup data is already available from the time the data is initially ingested, providing cohesive and consistent support throughout the entire lifecycle, including its feedback loops.

Hence, next-generation metamodels facilitate the continuous setup and dynamic evolution of an organization's metadata. This is distinct from activated metadata, which primarily aggregates business metadata mapped to the metadata about existing systems.

Employing metagraph technology, these dynamic metamodels assist business experts in capturing important contexts and complex dependencies beyond the capabilities of any single system. Moreover, they enable direct collaboration among knowledge workers, enhancing their collaboration and ability to manage complex daily operations.

The Setup Data Landscape

Types of Setup data: Master data, Reference data and Meta-data

Business information systems have primarily been designed to support business operations within and across different business functions throughout the value chain. Over time, organizations have become more adept at integrating these operational systems with shared data and unique identifiers, referred to as master data, about entities such as customers, suppliers, products and assets.

This approach to store and govern data about shared entities in one place has been very important to enhance reliability and traceability within and across business operational systems.

Reference data is commonly represented as a distinct subset within master data systems, used for categorization of data throughout a system landscape. It defines the characteristics of master data entities as well as the transactions and activities they are involved in. Reference data can be viewed as the simplest form of describing other data and as a basic form of metadata. Now, increasingly powerful tools are available to handle descriptive and defining contexts. The ability to classify, compose, and categorize entities and their dependencies is referred to as the 'Level of Abstraction'.

The trend among master data solution providers is shifting towards the capability to connect multiple master data entities, known as multi-domain master data. This includes information such as which products can be purchased from specific suppliers and the geographic regions where these suppliers are located.

Although many master data solution providers claim to offer multi-domain platforms, in reality, most can handle various types of entities but often fail to connect them effectively. We define 'Grade of Connectivity' as the ability to explicitly express interdependencies between master data entities and metadata within the data landscape setup.

To excel in an increasingly complex world, organizations must significantly improve their ability to manage both abstraction and connectivity. They need to master the complex context in which they operate.

The greater the interdependencies between different entities that an organization must handle, the more crucial it becomes to abstract in order to understand, manage and communicate complexity. Any duplicated dependency described and maintained at a lower level than it truly exists poses a risk of inconsistency and misunderstanding.

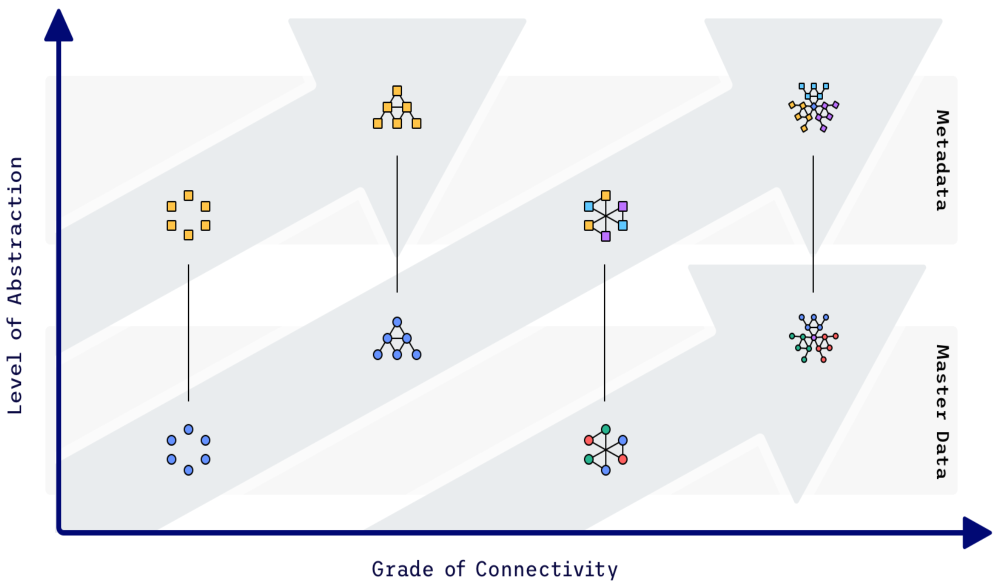


Figure 1. Evolve Setup Data Contextualization

Contextual Metadata

Contextual metadata encompasses various types of information that provide additional insight or background on a specific situation, event, or data point. Understanding this context is pivotal for interpreting data; it is essential for comprehension, decision-making, and interoperability. The success of digitalization strategies heavily depends on the ability to define and contextualize data from its origin.

Certainly, there are datasets that cannot be structured prior to registration; however, these are not as common as they are often portrayed. Moreover, by initially organizing the datasets that can be structured effectively, we will also create a stronger foundation for processing unstructured data, for example, through large language models (LLM).

In most cases, business operational systems such as Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), Product Lifecycle Management (PLM), and Supplier Relationship Management (SRM) excel at managing specific types of information at the operational level. However, they often struggle in expressing comprehensive interdependencies at more abstract levels. Introducing contextual metadata can bridge this gap and complement these operational systems by providing advanced configuration capabilities.

Contextual Metadata Supporting Knowledge Work

Business operational systems primarily manage large volumes of transactional data. The business context, however, encompasses all crucial information needed to define key aspects, such as which products should be sold, who manufactures and delivers them, and the methods, locations, and timing of these processes.

While operational systems process much of this information, coordination issues among various systems within the value chain often pose challenges. Additionally, business operational systems struggle to handle variations in products, services, production, and aftermarket needs. Consequently, these systems require reinforcement through a stronger contextual framework and enhanced capabilities for abstraction and continuous configuration to adapt swiftly and flexibly to increasing complexity.

Contextual metadata is essential for supporting the overall structure and understanding of an organization's landscape and all its generated data. This metadata provides the necessary insights and background information needed to make sense of high-volume data, including transactional data, sensor data, and natural language sources.

The following passage describes the relationships among various types of setup data, including master data, reference data, and metadata. It also explores how tools that are increasingly utilized, such as ontologies and knowledge graphs, connect to contextual metadata.

Additionally, it explains how these tools are emerging to meet the growing demand for well-structured knowledge management and analytics. This capability is crucial for enhancing the management of complexity throughout the entire lifecycle of a business, including its products and services.

The Nature of Setup Data and the Need of Metamodels

Data management refers to the process of collecting, storing, organizing, and maintaining data in a manner that ensures its accuracy, accessibility, reliability, and security throughout its lifecycle. Although data management naturally encompasses all data within an organization, managing setup data is particularly crucial. This paper aims to explain various types of setup data using a simple framework, providing clarity on their capabilities across different business levels.

The setup data can roughly be divided into four categories. On the x-axis, we have the model's or system's ability to connect multiple entities, referred to as the '*Grade of Connectivity*.' On the y-axis, we have the model's or system's ability to zoom out and reflect a greater whole, referred to as the '*Level of Abstraction*.'

This categorization results in the following four types of setup data:

- Single-domain Master Data
- Multi-domain Master Data
- Single-domain Metadata
- Multi-domain Metadata

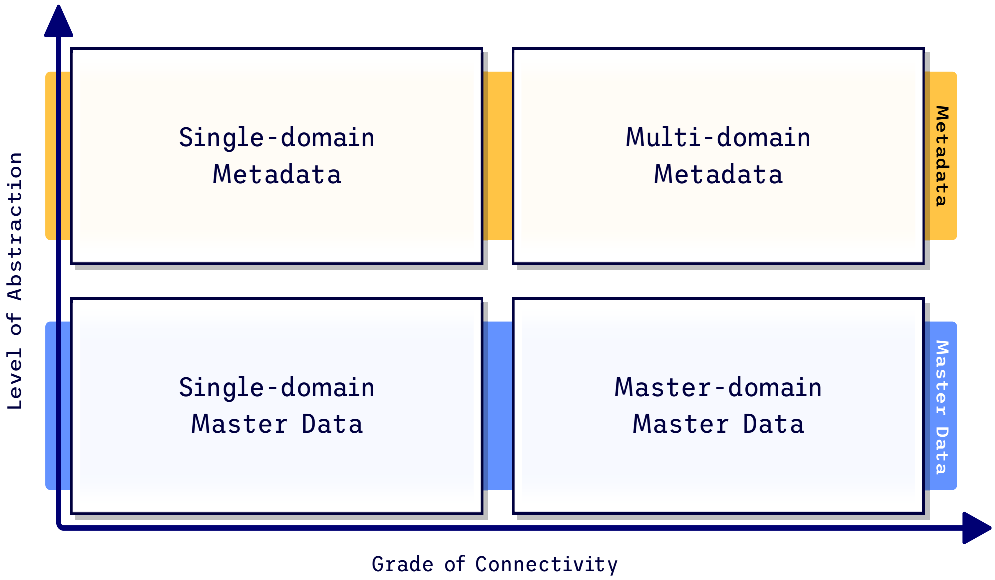


Figure 2. Four-Field Categorization of Setup Data

What is Master Data?

Master data has played a foundational role in the success of organizations implementing large-scale information systems by ensuring data consistency, quality, integration and operational efficiency. Master data refers to the core data entities that are essential to the operations of a business. It represents the foundational entities that are shared across various information systems and processes within an enterprise

> *"Master data is the consistent and uniform set of identifiers and extended attributes that describes the core entities of the enterprise including customers, prospects, citizens, suppliers, sites, hierarchies and chart of accounts."*
> – Gartner

The centralization of storage and governance for data and identifiers shared across the organization is crucial for improving quality and traceability within and between business systems. Master Data Management (MDM) is key in this process, overseeing the creation, maintenance, and governance of master data to ensure its accuracy, consistency, and reliability across the organization.

This is vital for maintaining data integrity and supporting quality data in decision-making.

What is Single-domain Master Data?

Single-domain master data refers to master data that is not linked to other master data domains. It meets the fundamental requirement of master data by ensuring traceability and reliability of individual entities within and across an organization's operational systems.

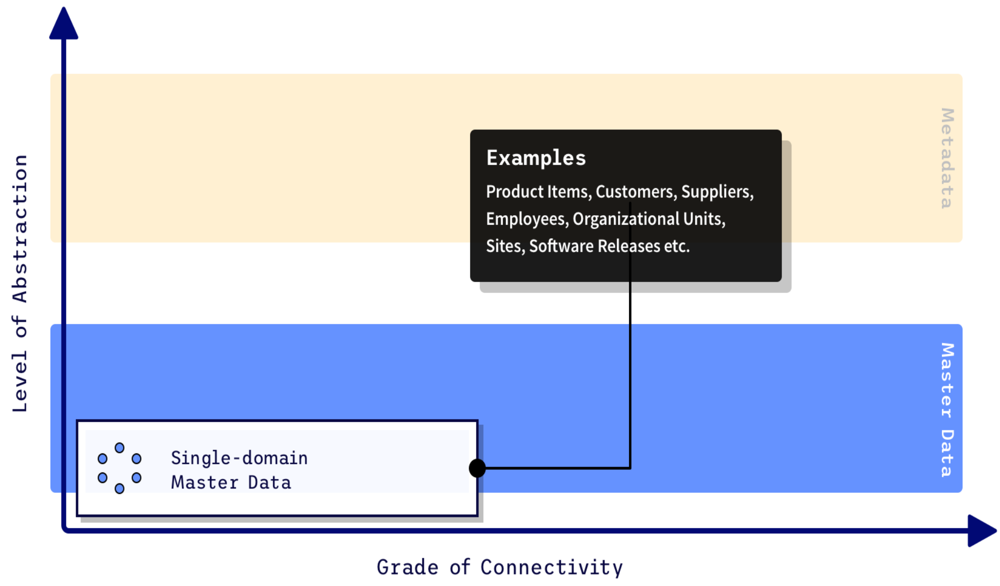


Figure 3. Master Data Positioning

In single-domain master data management, organizations focus on ensuring the accuracy, consistency, and integrity of data within specific, individual domains. It's important to note that, in this framework, even a system capable of handling multiple domains is still classified as a single-domain master data system if it cannot interconnect these domains.

What is Single-domain Reference Data?

Reference data acts as foundational data used to categorize, classify, and describe other data. It provides a framework for organizing information and ensuring consistency across systems and processes. Reference data typically changes slowly and serves as a basis for data analysis, reporting, and decision-making.

Moreover, reference data usually encompasses smaller volumes of data compared to master data. Single-domain reference data refers to standalone reference data that is not connected to any other reference data or specific context.

Typically, attributes used to characterize master data are also considered master data. However, much of the descriptive data for a master data entity can advantageously be treated as reference data. For instance, attributes describing a product's characteristics or functions have not traditionally been considered reference data.

With the advent of new metagraph technology, there are significant opportunities to treat a product's attributes and functions as metadata to be referred to from the entire system landscape. This enables consistent product descriptions to be used as a coordinate system across product lines and at all levels of abstraction, such as Product Type, Product Line, Product Variant, Product Item, and Product Individual.

Attributes and their corresponding product types can be inherited and instantiated from higher to lower abstraction levels, extending down to specific products. Additionally, treating each attribute as reference data with its own unique identifier facilitates consistent management of product data at the appropriate abstraction levels throughout the product lifecycle. This approach can also be applied to other master entities, such as customers, suppliers, and assets.

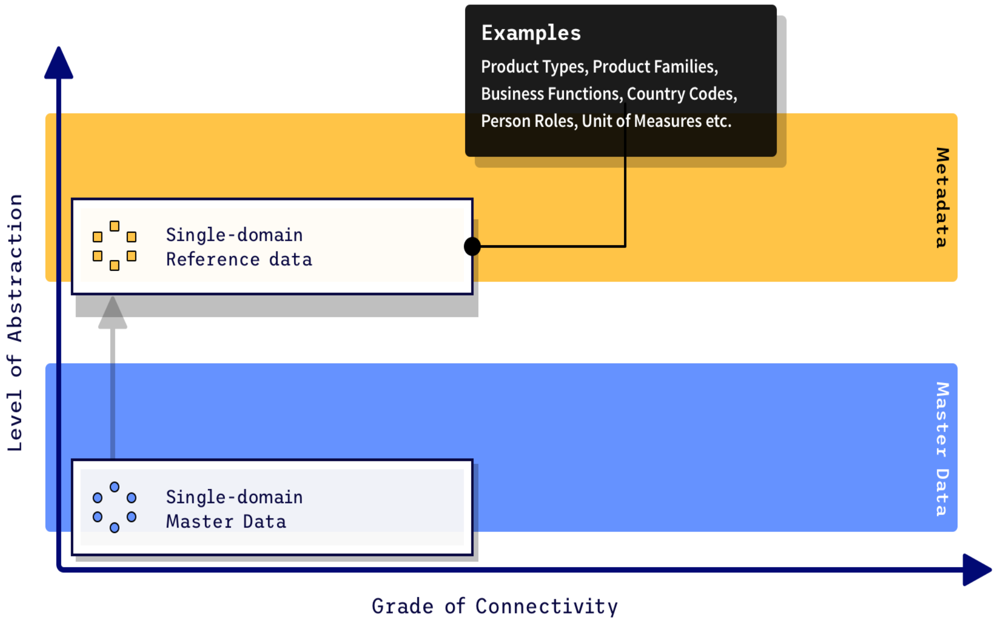


Figure 4. Reference Data Positioning

Oftentimes, reference data is considered a special subset of master data and is included within master data systems. However, in this graphical representation, we view reference data as metadata, in its simplest form, which means it is data about data.

What is Single-domain Structured Master Data?

We define single-domain structured master data as data and dependencies that describe the internal structure of an entity. Examples include product bills-of-materials (BOMs), supplier organization unit structures, internal organization unit structures, customer organization unit structures, and software unit installation breakdown structures.

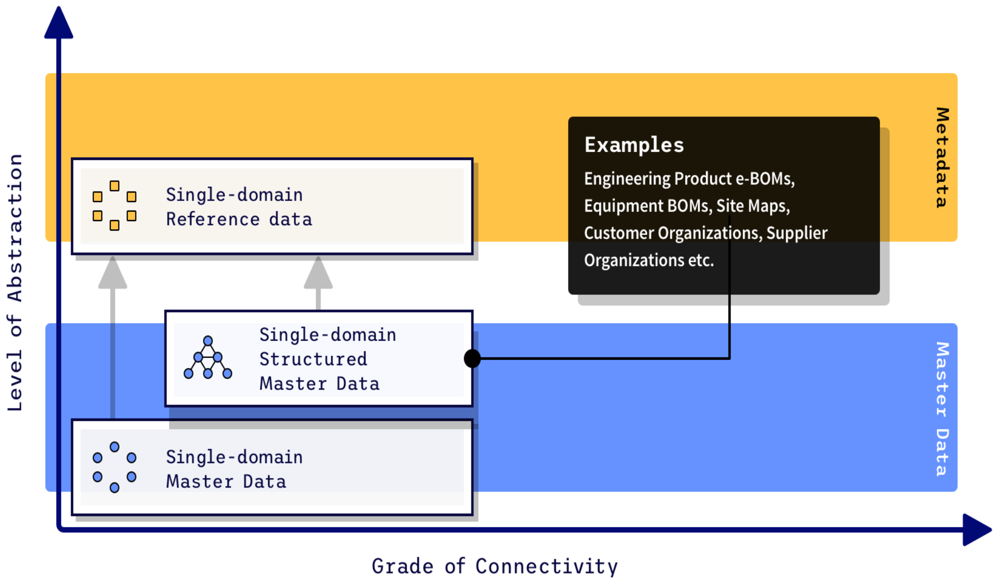


Figure 5. Single-domain Structured Master Data

Single-domain structured master data facilitates an understanding of how different parts of an entity are related. Typically, this data is organized hierarchically, as seen in a product's bill of materials. It can also cover multiple aspects within the same domain, exemplified by a matrix organizational structure.

What is Multi-domain Master Data?

We define multi-domain master data as data that describes the relationships between different master data entities. This includes specifying which suppliers sell which articles, identifying who is responsible for particular articles, and detailing the localization of retailers or the placement of installed products at customer sites.

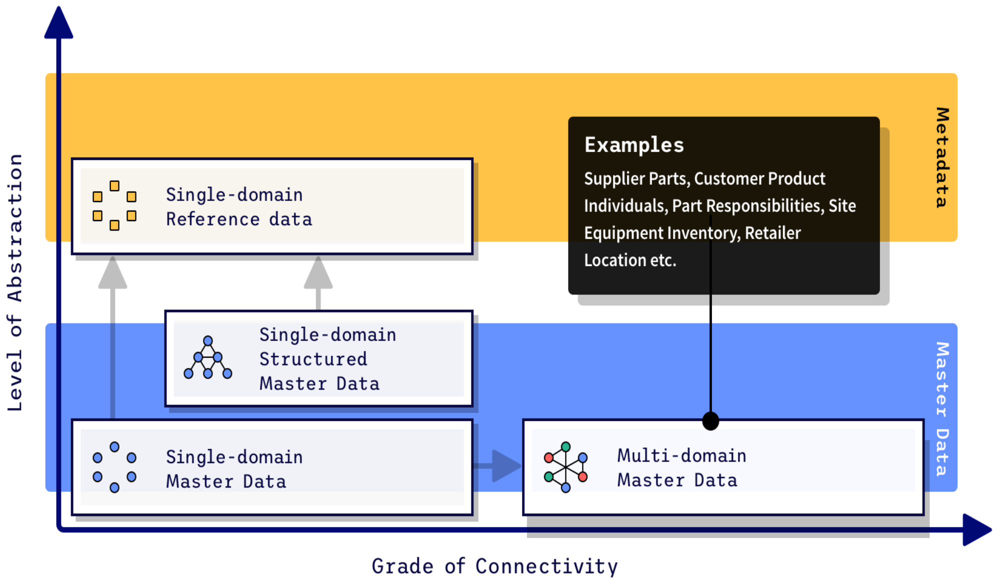


Figure 6. Multidomain Master Data

Connecting master data domains and tracking relationships that are common and important across the organization naturally offers high value. However, to achieve understanding and interoperability, the connected master data must be associated with relevant contextual metadata.

What is Multi-domain Structured Master Data?

Multi-domain structured master data involves connecting master data entities within a structured framework across different domains. This allows for the maintenance of simple relationships between specific data sets in a single location within the system landscape.

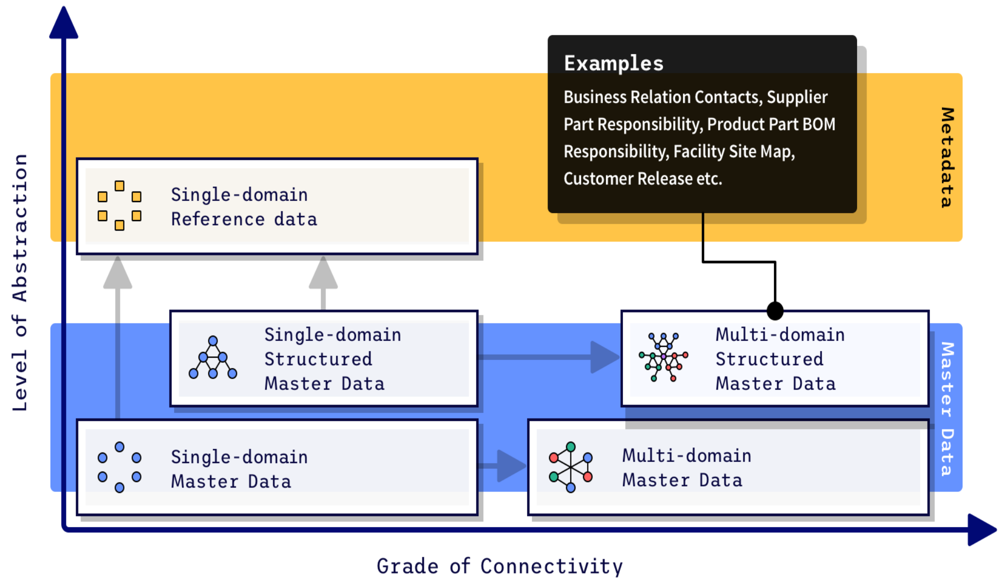


Figure 7. Multi-domain Structured Master Data

For example, multi-domain structured master data enables the determination of who within an organizational structure is responsible for specific products within a bill-of-materials.

What is Metadata?

Metadata refers to data that describes other data, providing information about a data item's content, attributes, or characteristics. It serves various purposes in different situations, such as describing content and its context, facilitating search, and aiding in data management. Metadata includes descriptive information about diverse aspects of a business, including its products, services, processes, competences, assets, customers and information systems.

The greatest challenge in digital transformations lies in the fact that metadata for business content is often tightly locked within specific system solutions. This restricts organizations' ability to adapt to change and manage increasing complexity. The presence of many different systems, each with its own semantics and method of storing metadata about business operations, leads to a fragmented toolchain characterized by inconsistent definitions of metadata.

There are now significant infrastructure changes underway to address this major problem: the active metadata approach and the dynamic metadata approach. The active metadata approach, which is gaining traction, helps interpret, transfer, and align information from diverse data sources within an organization. This metadata, also known as activated metadata, reflects business operations as recorded in various systems and database structures from multiple datasets.

Unlike previous technologies, in which metadata was essentially hidden within systems and databases, active metadata offers the flexibility to adapt to additions and changes across all data sources.

The dynamic metadata approach is enabled by the new metagraph technology, which is designed to manage complex contexts throughout a business's lifecycle. Dynamic metadata not only serves as a master reference for more specific data but also functions as a dynamic template.

This template assists in continuously designing, configuring, simulating, and realizing the business, including its products and services. It provides the tools necessary to transition from being solely data-driven to being both model- and data-driven, supported by a continuous and aligned feedback loop.

Fundamentally, it's not about choosing one approach over the other, but about balancing their usage efficiently over time. The active metadata approach is powerful for gaining an understanding of the data handled by existing systems and data sources in the short term. It facilitates insights into what has happened in the business through the lens of its existing information systems.

In contrary, the dynamic metadata approach supports the entire lifecycle, including operations and analytics, and is naturally the long-term direction to take. This approach creates entirely different conditions for knowledge workers by enabling them to receive direct feedback on their activities and decisions in operational tasks, all in a unified business language. It also offers unique opportunities to utilize AI and machine learning based on well-defined and structured data that can be instantly supplied in the process flow.

The activated metadata approach is effective until you have gradually begun to establish the foundation for dynamic metadata. However, keep in mind that investing in dynamic metadata, prior to activated metadata, will significantly enhance your contextual capabilities and support for knowledge and data workers, both in the short and long term.

Contextual Metadata, Ontologies and Knowledge Graphs

There are several methods for modeling contextual metadata and its associated content, with knowledge graphs and ontologies being two primary approaches used in data science and AI. These techniques are valuable for enhancing the understanding and analysis of business data, often utilizing methods such as data mining.

Ontologies tend to focus on the conceptualization of a domain, defining types, properties, and interrelationships among entities that fundamentally exist within that domain. In contrast, knowledge graphs tend to represent specific instances of those entities and their relationships.

There is a debate regarding whether knowledge graphs and ontologies should be more closely interrelated or even considered as the same thing, centering on the continuum from abstract conceptualization to concrete representation of knowledge.

Together, these concepts represent a layered approach to organizing, interpreting, and acting upon knowledge. Ontologies, including taxonomies, define the conceptual framework, knowledge graphs fill the framework with interconnected data, and metadata provides the means to dynamically interact with and utilize this data.

Dynamic metadata, enabled through metagraph technology, facilitates the alignment of ontologies with knowledge graphs within a single, consistent structure. In the realm of AI, it has become increasingly evident that relying solely on traditional knowledge graphs is insufficient for achieving reliable outcomes. For instance, in the context of Large Language Models (LLMs), the significant impact of a well-structured approach on the quality of results is clearly evident.

Given that current knowledge graphs, which rely on conventional graph technology, no longer provide sufficient results for advanced AI applications, it's time for progress. The knowledge graph community is now shifting towards hypergraphs, which can represent more complex relationships by connecting more than two nodes. Hypergraphs are an extension of graph theory, where each edge, known as a hyperedge, can connect more than two nodes.

In a standard graph, an edge connects exactly two nodes, but in a hypergraph, a single hyperedge can connect any number of nodes. This makes hypergraphs a powerful tool for modeling relationships not limited to pairs, allowing for a more truthful representation of complex contextual dependencies.

Linking specific, instance-level information with higher-level, abstract definitions is central to our contextual approach to setup data. Thus, it is crucial to consider knowledge graphs and ontologies as a unified entity that enables the extraction of knowledge across various levels of abstraction. Metagraph technology supports us in tracking complex interdependencies at all levels of abstraction.

We refer to this ability to model contextual metadata using the metagraph technology as '*meta-ontology*' or '*knowledge metagraph*' depending on the background of our audience.

What is Multi-domain Metadata?

Multi-domain metadata refers to the ability to link metadata from one domain with metadata from other domains. This facilitates the development and evolution of metadata contexts by, for example, associating products with services or linking services with resources at a general metadata level.

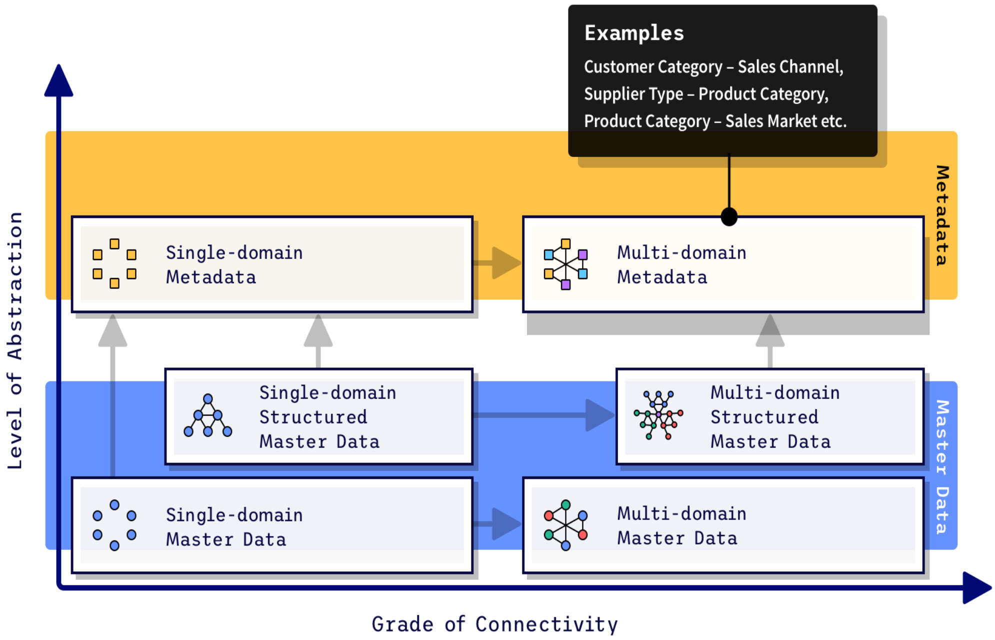


Figure 8. Multi-domain Metadata

What is Single-domain Structured Metadata?

Single-domain structured metadata represents a significant advancement over Single-domain Metadata. It facilitates the construction of structured entities within metadata. It is crucial to distinguish between compositional structures and classification structures. A typical example of a compositional structure is a Bill of Materials (BOM), which provides a generic breakdown, such as that of a bicycle. This serves as a foundational structure, acting as a backbone for potential variations and desired configurations. Other examples of breakdown structures include generic software architectures and organizational breakdowns.

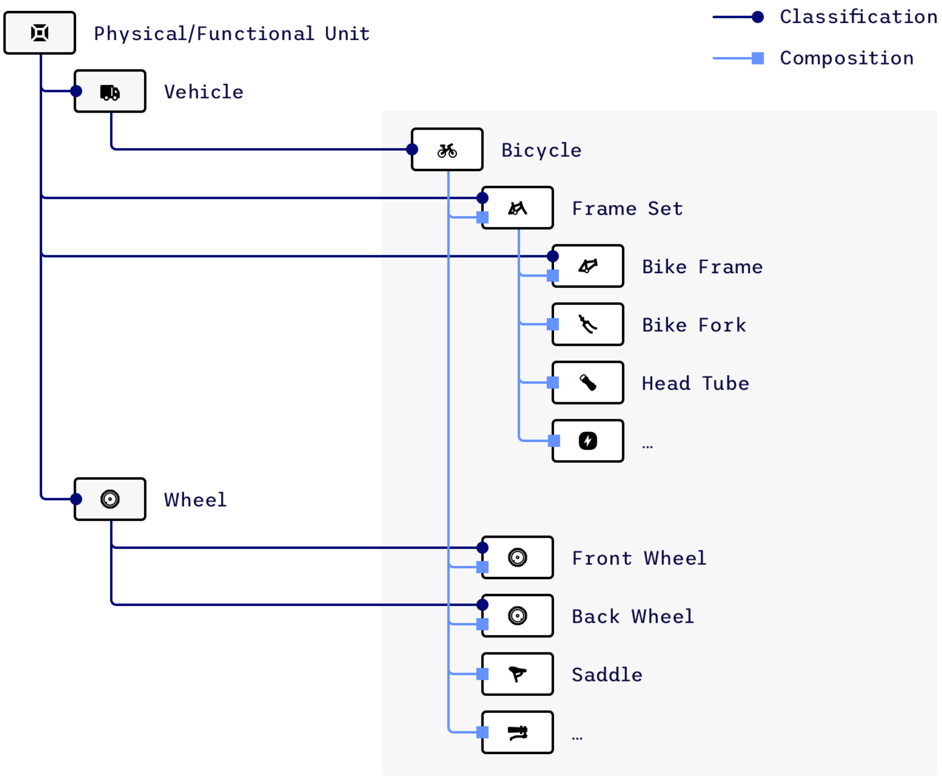


Figure 9: Classification and Composition Structure

An example of a classification structure, also referred to as a taxonomy, can be found in the classification of different kinds of bicycles: mountain bike, road bike, and hybrid bike. The mountain bike, in turn, can be subclassed into a cross-country bike and a trail bike, and so on.

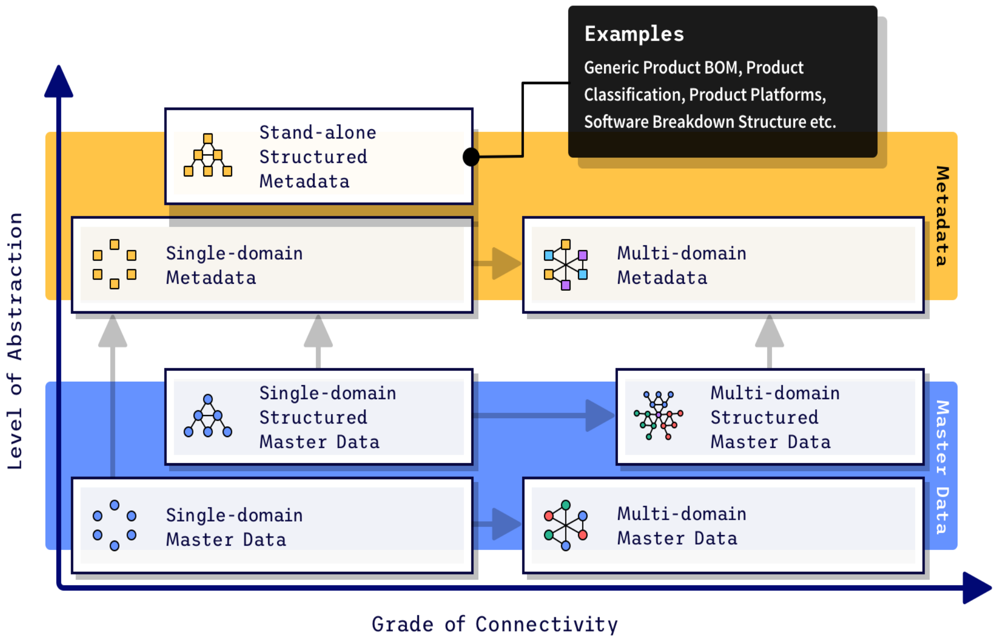


Figure 10. Single-domain Structured Metadata

Solutions are now available to manage various types of relationships among different disciplines within product structures. This is where new metagraph technology proves useful, not only for describing hierarchical BOM (Bill of Materials) structures but also for illustrating cross-disciplinary relationships, for example between software, electronics, and hardware. Combined with well-developed abstraction capabilities, this is crucial for handling the significant variances that arise in these increasingly complex systems moving forward.

What is Multi-domain Structured Metadata?

Multi-domain structured metadata enables the description of complex relationships at all essential abstraction levels within a business, playing a crucial role in continuously adapting to the ever-changing conditions of complex businesses.

Additionally, this facilitates interoperability and collaboration beyond the capabilities of existing information systems and established silos. It is essential for managing the design, configuration, simulation, realization, and lifecycle management of businesses, including their products and service.

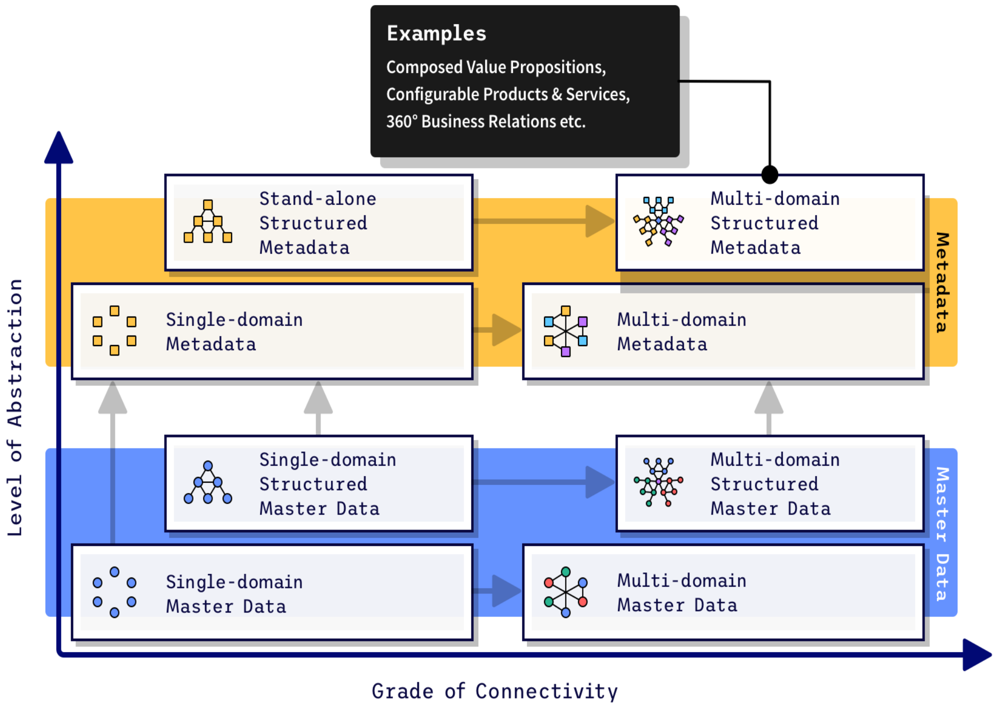


Figure 11 Multi-domain Structured Metadata

Examples of multi-domain structured metadata cases include multi-discipline product design and configuration, product and service portfolios, 360-degree business relationship management, digital twin lifecycle management, and metagraph-driven AI and LLM capabilities.

In summarizing the setup data framework, we use the terms ‘contextual metadata’ and ‘meta-models’ interchangeably across all metadata categories, except for single-domain metadata. We generally employ ‘contextual metadata’ when discussing from a data perspective to highlight the importance of considering it within its context. Similarly, we use ‘metamodel’ to underscore the importance of designing and developing a model that serves as a foundation for thoughtfully and systematically structuring metadata.

We also use the concept of a metagraph, which is a metamodel implemented using metagraph technology. Similarly, we use the term ‘meta-ontology’ to describe ontologies and the term ‘knowledge metagraph’ for knowledge graphs that are developed through metagraph technology.

Furthermore, we use the concept of a sustainable metamodel for metamodels that can be expanded to handle additional levels of abstraction and multidimensional relationships.

Sustainable meta-models can be expanded to meet the needs of an entire organization, adhering to the principle of a single source of truth.

A Solid Setup Data Infrastructure for the Entire Business

Developing the ability to handle contextual master data with the support of metamodels and metagraph technology does not negate the importance of considering master data and its relationships at the specific level. The goal is to manage master data, metadata, and supporting metagraph models effectively in a comprehensive manner, enabling the business to track all complex dependencies across operational, tactical, and strategic levels

Figure 12 Setup Data Mastery

Master data is essential for achieving traceability and reliability at the operational level of a business. Moreover, reference data and more advanced multi-domain metadata play a key role in enhancing understanding and transparency at a higher operational level. Equally important, multidimensional metadata, supported by highly abstract metamodels, is vital for gaining insights and determining the business's strategic and tactical positioning.

These capabilities are crucial for navigating complex businesses with intricate product and service offerings. Developing a map that outlines the requirements for set up data across the entire business not only creates the conditions for properly managing business-critical metadata but also places master data within a broader context.

This fosters entirely new conditions for organizations and business experts in the field to develop a shared understanding of and engagement with the continuous and dynamic changes in both metadata and master data.

Setup Data Solutions; From Master Data and Metadata to Metamodel Mastery

A basic understanding of setting up data, with its various complexities and at different levels of abstraction, is crucial for establishing a sustainable information and system infrastructure across the entire organization.

To select the best information systems for managing setup data across various levels of abstraction and connectivity, it is important to thoroughly understand the specific needs of the business and the capabilities of different systems.

By effectively contextualizing master data with metadata, we can involve key knowledge workers from various parts of the organization in the management of setup data. It is crucial to keep these experts engaged by consistently providing strong support for their complex daily tasks.

Master Data Solutions

Master data solutions encompass a variety of software and platforms designed to manage an organization's critical entity data assets, ensuring consistency, accuracy, and accessibility across various operational systems. Generally, there is a risk in implementing master data systems without first having a good understanding of the larger context of the setup data.

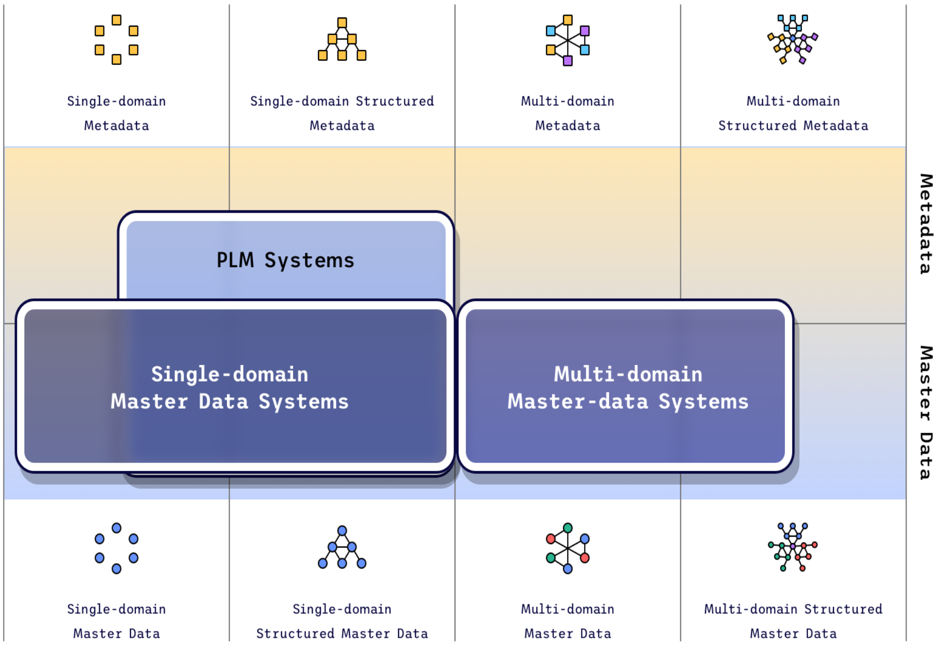


Figure 13 Master Data Management Systems

To implement master data solutions can be highly complex due to the need to integrate various data sources. This often requires significant IT resources and coordination across different departments, which can lead to extended timelines and higher costs. Changing organizational processes to adapt to a new master data system can be challenging.

Resistance from users accustomed to older systems and processes can hinder the successful adoption of the new system. Overall, many challenges in implementing master data systems arise because business experts feel burdened by an additional task that appears to add no direct value and only slows down their processes.

Although it remains a challenging task, the key to success lies in creating a common understanding that setup data is an essential component of the value-creation system, just as the correct parts must be assembled at the right time during physical production.

Metamodel and Metagraph Solutions

Like other complex constructions, it is crucial to approach information infrastructures and data strategies from a holistic perspective. A business functions as a value-creating system. Companies that develop and produce physical products often have a clear understanding of the structure of the value system within the value chain. However, as these companies increasingly produce digital products that mirror their physical counterparts, they often struggle to adequately understand and describe the value-creating system.

To handle this multi-dimensional complexity at various levels of abstraction, metagraphs and reference model solutions are now available. Similarly, just as master data solutions store master data in a repository, metadata solutions store metadata in a repository known as a Definition Master. Maintaining metagraphs in a Definition Master offers significant opportunities to develop aligned well-structured knowledge graphs and ontologies at the metagraph level.

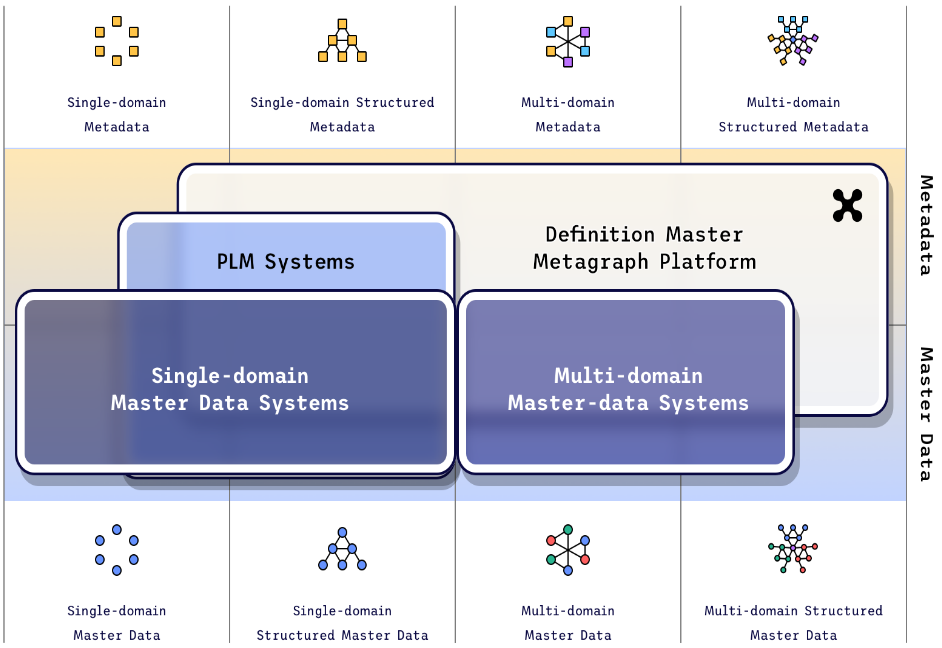


Figure 14. inorigo® Metagraph Platform Positioning

Conclusions

Effectively managing setup data and master data is critical for organizations seeking successful digital transformation. The key challenges involve engaging the right stakeholders and maintaining a balance between data quality and operational flexibility. Building a robust infrastructure that leverages meta-models and meta-ontologies can facilitate clear specifications of digital products and enhance the creation of digital tools within the value chain.

For organizations dealing with complex products and services, developing contextual meta-models that accommodate extensive configurations is essential. Moreover, the advent of knowledge metagraphs promises to revolutionize the integration of large language models (LLMs) into daily business operations, paving the way for more reliable and instantly served artificial intelligence systems.

All things considered, a well-structured infrastructure for set up data, provides completely new conditions for interoperability, between both systems and knowledge workers. Furthermore, it provides entirely new opportunities to zoom in and out of a business since identifiers at the master data level are well integrated with their definitions across all essential layers of abstractions.