Vysoká škola ekonomická v Praze

Fakulta informatiky a statistiky

Katedra informačních technologií

Studijní program: Aplikovaná informatika

Obor: Informační systémy a technologie

MDM of Product Data

DIPLOMOVÁ PRÁCE

Student : Bc. Lenka Čvančarová

Vedoucí : doc. Ing. Jan Pour, CSc.

Oponent : To be defined

2012

Prohlášení

Prohlašuji, že jsem diplomovou práci zpracovala samostatně a že jsem uvedla všechny použité prameny a literaturu, ze které jsem čerpala.

V Praze dne 12. prosince 2012 ……..

Jméno a příjmení studenta

Abstrakt

Tato práce se zaměřuje na Master Data Management produktových dat. V současné době se většina publikací o MDM zaměřuje především na data o zákaznících a pouze velmi málo zdrojů popisuje problematiku produktových dat. I přesto, že se některé zdroje pokoušejí popsat MDM v celkovém pohledu, ve většině případů se nakonec zaměří právě na zákaznická data.

Nedostatek zdrojů o MDM produktových dat se stal jedním z motivací pro tuto práci. Dalším důvodem bylo nastínění a analýza specifik MDM produktových dat v kontextu implementace a požadavků na aplikační software z hlediska firmy, která se zabývá vývojem aplikací pro MDM.

Z tohoto důvodu jsem vytvořila a nastínila metodiku pro zavádění MDM produktových dat. Ta byla vytvořena na základě osobních zkušeností z MDM projektů se zákaznickými daty, které jsem aplikovala na teoretickou část této práce. Analýza vlastností produktový dat a jejich dopadu na zavedení MDM a analýza požadavků na aplikační software v této práci umožní dodavatelům MDM řešení, kteří se zaměřují na zákaznická data, lépe porozumět problematice MDM produktových dat a následovně rozšířit své zaměření.

Kromě toho může tato práce sloužit jako zdroj informací pro společnosti, které rozvažují zavedení MDM produktových dat do jejich podnikové infrastruktury.

Klíčová slova

Master Data Management (MDM), Product Information Management (PIM), Customer Data Integration (CDI), Master Data, datová kvalita.

Abstract

This thesis is focused on Master Data Management of Product Data. At present, most publications on the topic of MDM take into account customer data, and a very limited number of sources focus solely on product data. Some resources actually do attempt to cover MDM in full-depth. Even those publications are typically are very customer oriented.

The lack of Product MDM oriented literature became one of the motivations for this thesis. Another motivation was to outline and analyze specifics of Product MDM in context of its implementation and software requirements for a vendor of MDM application software.

For this I chose to create and describe a methodology for implementing MDM of product data. The methodology was derived from personal experience on projects focused on MDM of customer data, which was applied on findings from the theoretical part of this thesis. By analyzing product data characteristics and their impacts on MDM implementation as well as their requirements for application software, this thesis helps vendors of Customer MDM to understand the challenges of Product MDM and therefore to embark onto the product data MDM domain.

Moreover this thesis can also serve as an information resource for enterprises considering adopting MDM of product data into their infrastructure.

Key words

Master Data Management (MDM), Product Information Management (PIM), Customer Data Integration (CDI), Master Data, Data Quality.

Table of Contents

1 Introduction

Data and information are crucial for forming any enterprise strategy. Therefore most enterprises have multiple repositories for key business data (customers, suppliers, products, etc.). Thanks to the great impact of analytical activities within the enterprise, as well as on decision making, consistent, good quality data, and their convenient management, is a basic requirement for each company's success.

The rise in the adoption of enterprise applications (such as ERP, SCM and other systems) in the past decade has turned the control of accurate and consistent data into a challenge regarding the many places in which the data is stored. Master Data Management (MDM) plays a significant role in reducing data discrepancies and enabling central view of data across the enterprise in order to keep data centrally managed, updated, consistent and without duplicities.

Product information is a fast changing asset, and therefore, it must be referenced and managed efficiently, accurately and in an up-to-date fashion in order to keep up with competitors. Product data flows through the enterprise and the supply chain, from customers, through retailers, and eventually to raw material suppliers. With regards to the current market trends, such as speeding up a product life-cycle, speed and number of products introduced or terminated, company merges, acquisitions, expansions to foreign countries, and a global network of suppliers, these all lead to problems with managing product data.

Many resources emphasize the importance of Master Data Management of Customer as it is according to TDWI[1], by far, the entity mostly defined via MDM [RUSS2012]. This dominance is caused by enterprises traditionally trying to better understand and serve their customers; though in product oriented industries the importance is put on product data.

According to Gartner[2] the Product MDM exceeded the Customer MDM in 2011 in terms of software spending [GAR2011]. The reason behind may be because manufacturing and retail industries are catching up at automating information technology. Therefore the product segment of MDM has started to rise and has a great potential after the need for Customer MDM has been satisfied.

[1] TDWI (The Data Warehousing Institute) is an educational institute of BI and data warehousing.
[2] Gartner is a technology consulting firm.

1.1 Aim

The aim of this thesis is to analyze product data characteristics, requirements and their impact on activities covering implementing MDM of product data into the enterprise in order to create a methodology for successful implementation of Product MDM solutions.

Its individual aims cover:

- Analyzing product data characteristics and requirements.
- Classifying and analyzing sectors for MDM in product data usage and point out their unique industry-specific requirements and benefits.
- Providing an overview of the current market for Product MDM application software.
- Identifying requirements for Product MDM application software.

My aims do not cover:

- Providing metrics measuring business improvement and benefits of implementing Product MDM.
- Analyzing solution integration into the company infrastructure and architecture.
- Assigning internal or external roles (neither teams nor individuals) to tasks connected with implementing the solution into the enterprise.

As there were no product-data examples present for the use of this thesis, no real life product data could have been used for the analysis, and a possible use case from an enterprise. Therefore this thesis intends to research the segment of Product MDM in order to create a knowledge base for future PIM projects.

1.2 Solution Approach

This thesis consists of two main logical units.

The theoretical unit is focused on concepts and benefits of master data management and product data characteristics. The analysis is based on various sources focused of MDM and product data. This builds up the required theoretical background for the second unit.

The practical unit of this thesis contains a methodology for adopting an MDM solution of product data, which is based on the findings within the theoretical unit. The second unit consists of listing

requirements for a software tool derived from product data characteristics, analysis of specific enterprise needs based on it processes and industry and major steps and activities in adopting MDM of product data, which is also based on my work experience.

1.3 Thesis Structure

First chapter is a general introduction to the topic. It contains the aim, structure and approach to the problem as it is perceived in this thesis.

Second chapter provides an overview of different approaches and definitions of MDM as it is understood by some of the authorities in the field. It also outlines a brief description of MDM benefits and stresses the importance of master data, reference data and metadata in an enterprise. It also introduces a general concept of a domain in MDM as well as the most common types of MDM domains.

Third chapter focuses on product data and Product Information Management (PIM). It presents unique characteristics and requirements of product data as well as different ways of how they influence approaching and managing product data. It also elaborates on MDM benefits in greater detail in context of product data and its processes. An overview of the PIM market and its software tools used for MDM of product data closes this chapter.

Fourth chapter contains the actual methodology of PIM adoption. It is split into three logical subparts each covering one important aspect. First it focuses on unique requirements for a software tool regarding MDM of product data. Second it emphasizes the company background and the influence of its processes and industry sector on PIM adoption. At last it lays out essential processes taking place while adopting PIM emphasizing characteristics and specific needs of product data.

Fifth chapter is a conclusion of this thesis.

2 Master Data Management

Managing key business data, be it customer data, supplier data, etc. has always been one of the main concerns of every enterprise, no matter the industry involved.

First attempts for MDM arose in the wake of ERP boom in the 1990s [GLA2000]. In this period, enterprises were buying, or building enormous enterprise applications; providing managers with a wealth of data. Each application had its own records with a specific list of attributes. As functionalities were becoming more and more complex, enterprises needed to spread their data across many other systems and applications as each of them required different data sets. As a result, these application-specific attributes were either added to the present records, or saved in as a different database record. This resulted in complex sets of data scattered in different places within the enterprise. MDM provides a central repository for all applications without the need of point to point integration among individual data sources. This allows easier management of disparate data.

Nowadays, enterprises are facing other challenges which are difficult to handle without MDM. As enterprises expand into foreign markets, globalize, merge and make acquisitions, it is also necessary to centralize and manage their data in greater scale. MDM enables them to merge their data into a central repository, find duplicities and standardize it into consistent sets of attributes which may be used across all departments.

This chapter describes the purpose of MDM and its use in organizations. For this purpose conclusions are used from a report on MDM conducted by TDWI (The Data Warehousing Institute) in 2012. This report is based on 369 survey respondents from all over the world [RUS2012].

2.1 Defining MDM

MDM aims to provide a company with a central repository which collects and propagates consistent, consolidated and standardized data across different IT systems and departments not only within the enterprise (like branches, warehouses) but possibly also among other external subjects (partners, customers and others) regardless of the solution technical approach.

This functionality implies technologies from system to data integration. However, the core processes of the whole MDM solution involve data operations, such as data cleansing, enriching,

validating, scoring, and matching duplicate entries, which are then fed into enterprise applications. Definitions of master data management used in this thesis below are based off of those from the following institutions:

(1) Adastra[3] [NET_02]
(2) TDWI (The Data Warehousing Institute)[4] [RUS2012]

The description of the term "MDM" varies from a strategy to a set of processes. Generally it is understood as an approach towards maintaining and distributing key data within an organization. Some institutions perceive it as mainly data quality oriented, other institutions focus on the business process point of view. Regardless the definition used, it is always considered as an ongoing initiative with the implementation being just the first milestone.

Definition by Adastra:

"Master data management (MDM) comprises a synoptic set of business proceedings (processes, policies and tools) which centrally and persistently define the key data of an organization (also called Reference data). The objective of MDM is to collect from, and supply to various processes, unique instances of each entity (single version of the truth). Uniqueness in this context is an aspect of data normalization which eliminates outright redundancy due to collection of instances from dissimilar sources. MDM ensures the consistency, accuracy, stewardship and accountability for the core information of the enterprise, enabling organizations to eliminate endless debates about "who's data is right"." [NET_02]

Definition by TDWI:

"Master Data Management (MDM) is the practice of defining and maintaining consistent definitions of business entities (e.g., customer or product) and data about them across multiple IT systems and possibly beyond the enterprise to partnering businesses. MDM gets its name from the master and/or reference data through which consensus-driven entity definitions are usually expressed. An MDM solution provides shared and governed access to the uniquely identified entities of master data assets, so those enterprise assets can be applied broadly and consistently across an organization." [RUS2012]

[3] Adastra is a consulting firm specializing in data warehousing, MDM, data quality and BI
[4] TDWI is an educational institute of BI and data warehousing

2.2 Defining Data in MDM

Since data is the core element, it is necessary to describe and define data which have important roles within the scope of MDM.

2.2.1 Reference Data

In different sources the term “reference data” is defined as all of the three types below, sometimes as a single group, sometimes as only one type of the three. For the purposes of this thesis the following definitions are going to be used:

2.2.1.1 Internal Reference Data

The term internal reference data can be applied on any limited list of values which is usually company specific. A typical example is a list of all values used for defining marital status, or other attributes.

2.2.1.2 External Reference Data

External reference data typically is maintained by an external institution (government or a private company). It may be acquired either free-of-charge, or at specified costs. Typical examples are lists of:

- Countries (including ISO codes)
- Currencies (including currency names)
- Industry codes (such as NACE)
- Different product classification standards

2.2.1.3 Shared Reference Data

Same as described in the previous cases, reference data may also refer to the complete list of all customers, products, addresses, orders and other entities within the company, including all related attributes.

2.2.2 Master Data

Master data is a standardized, enriched and consolidated data set usually stored in a central master data repository. They capture key data that all parts and departments of an enterprise agree on in different uses. It is important that the whole enterprise has a unified understanding of what is the

definition of key entities (such as customer, supplier, and product) and relationships among them. This prevents inconsistent reports and actions within the enterprise. Master data is important in both operational and analytical environments:

- Operational systems – many business processes use master data, a typical example would be adding a new email to a customer, introducing a new product to the market. Such processes use various application systems, but the data they work with must be consistent and accurate.
- Analytical systems – they usually have similar requirements to the operational systems. Master data often forms the key dimensions and hierarchies used for reporting and key business data analysis, using the master data for reporting. Therefore trustworthy master data is equally important.

2.2.2.1 Internal Reference data

In different subsidiaries or departments of a company, there may be different internal reference data tables for the same attribute. A typical example would be several values expressing a gender of a person (for instance “F” and “1”, both having the meaning “female”). In order to keep consistency in the master table, such values are translated or standardized into the same set. An example of such table can be found on appendix 10.3 and 10.4.

2.2.3 Metadata

Metadata is a very broad term, mostly understood as a descriptive piece of information used for any kind of data. It can be used for tagging, organizing, providing description of files, tables or large amounts of data. It is commonly used to define legal data types for columns in databases and to maintain their structure. It is also used to tag unstructured data.

To manage an MDM solution, it is necessary to use metadata for describing how to use master data, their physical structure, relationships, references and other characteristics. It is because the same description or attribute name can be used for different values and sometimes attribute names can be ambiguous (e.g. product description, product ID, product code can contain the same data values because it depends how a person or enterprise understands it). This is when metadata is used to provide the data sets with context.

Metadata store information about data in MDM solutions to be understood and therefore well integrated and managed. In MDM solutions they are usually stored in specialized metadata repositories often visualized in models as objects with relationships.

2.3 Benefits

Each MDM solution is unique with a specific set of functionalities depending on the needs of the organization in question. Nevertheless there are several core benefits from MDM solution connected with the existence of shared and consistent master data which provides the best source of truth created from multiple disparate sources.

According to a study by TDWI [RUS2012] the top reasons for implementing an MDM solution are to enable complete views of key business entities (customers, products, employees, etc.) and share them consistently across the whole enterprise. Other reasons concern improvement of analytical and other data based decisions, operational excellence, and regulatory compliance.

Most MDM solutions support basic functions for offline aggregation and reference data standardization. Advanced functions, which depend on company needs, include real-time processing, two-way data synchronization and workflow and process management. As company MDM solution functions mature over time, more features can be added. As a result, vendor products evolve in order to better support customer needs, and MDM trends. [RUS2012]

Solutions to issues within MDM vary depending on the enterprise. Options consist of completely homegrown, to a dedicated software implementation by external vendors. The current state is that a majority of enterprises with an implemented solution has a homegrown MDM solution with only 12% having dedicated MDM software, as shown in a graph below. [RUS2012]

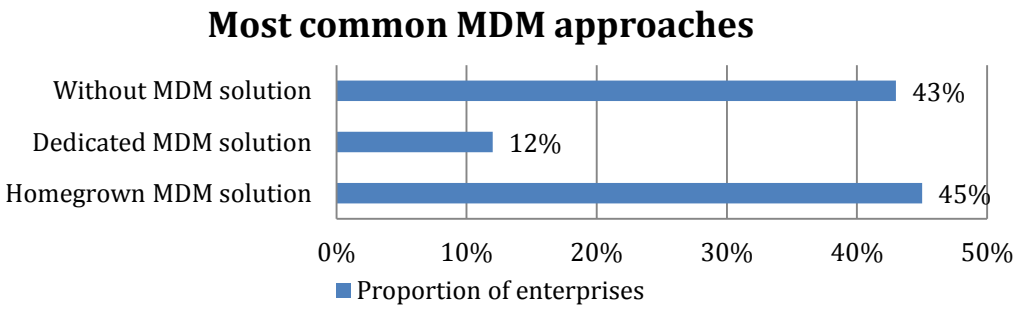


Figure 1 - Most common MDM approaches [RUS2012]

Benefits of MDM will be described in greater detail in relation to product data in chapter MDM of Product Data.

2.4 Domain Specific MDM

MDM domains are *"categories of master data, such as the customers, suppliers, products, and accounts. Each of these domains of master data represents information that is needed across different business processes, across organizational units, and between operational systems and decision support systems."* [DRE2008]

MDM may be applied to any kind of different data dimensions depending on the requirements of the company. Different types of data (such as party, product, location) require different approaches. This is why several main disciplines were established to distinguish product related MDM (Product Information Management), customer related MDM (Customer Data Integration), and other MDM types (most commonly locations, assets and accounts). Traditionally customer and product MDM solutions are the most frequently adopted solutions due to their strategic nature, and because many enterprises are in a customer focused industry.

According to a study by TDWI, the most modeled entity with master or reference data is Customers with 77%. Nevertheless the party-group oriented MDM solution, including Customers, Business Partners, Employees, Sales Contacts and Prospects, Patients and Citizens, are the most often adopted types. [RUS2012]

The graph below shows most modeled business entities in MDM. Respondents could choose multiple answers in case of several MDM solutions.

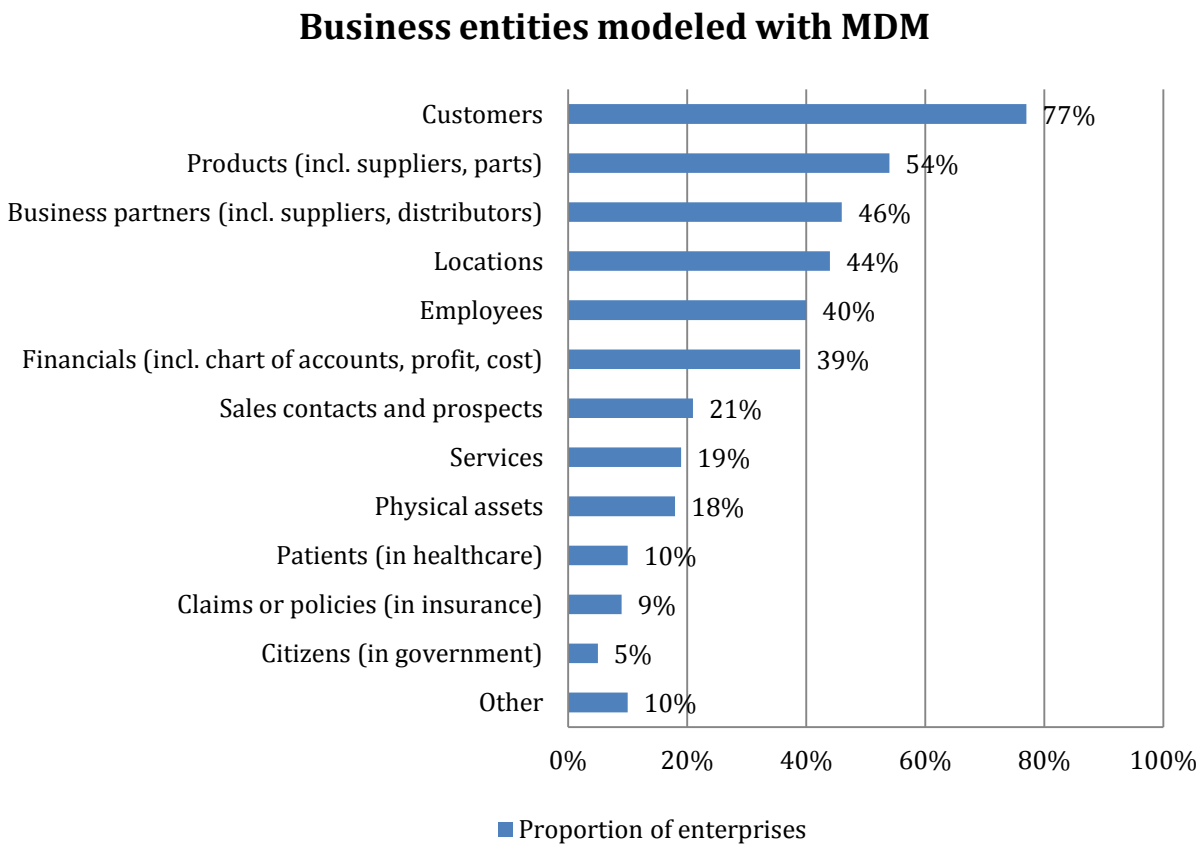


Figure 2 - Business entities modeled with MDM in enterprises [RUS2012]

The following sections cover basic information about MDM in most common domains (customer and location) and a basic concept of multi-domain MDM solution. Product data domain is thoroughly elaborated in the chapter MDM of Product Data.

2.4.1 MDM of Customer Data

Customer data integration (CDI) provides a single view[5] of the enterprise customers and its relations across the whole organization. There are many similarities between customer and citizen, patient, employee, suppliers, etc. including company data as well as it is of the same principle. This type of data is collectively called party data.

[5] Single view of customer/product/other entity is a term used to describe viewing an entity in MDM comprising all integrated data sources from any department of the enterprise.

2.4.2 MDM of Location Data

There are different ways to describe a physical location. It can be related as address data, territory, or latitude and longitude. Location MDM usually relates to specific type of data, e.g. geographic reference data or urbanistic data. Different types of industries may require geographic data (flooding, seismic, soil data etc.) in relation to several types of sociological data (e.g. population, ethnicity, income). Location data is most commonly related to either customer or other party data.

2.4.3 Multi-Domain MDM Solution

Multi-Domain (or multi-entity) MDM solution simultaneously manages several different master data domains across the enterprise.

Typically it is firstly implemented as a single domain solution (enterprises handle one type of master data, this is mostly either customer or product) which is added on one or more domains to their former MDM solution over time. Recently multi-domain MDM solutions have also been implemented to handle more than a single domain from inception without any MDM solution present in the enterprise at the moment of initial adoption.

According to Gartner [GAR2011] multi-domain solution are typically of the following characteristics:

- The data model is uniform or interoperable
- The data model can operate across multiple domains
- The workflow and user interface elements are uniform or interoperable
- It supports at least one use case, implementation style or organization/governance model, for specific industry scenario

2.5 Chapter Summary

Aims and definition of master data management are described in the introduction of the chapter. The chapter also contains definitions of different types of data present in MDM solution and their functions.

This chapter focuses on basic benefits of master data management which are described further in this text in relation to product domain. It also describes various areas of focus of MDM, related domain specific MDM types and their characteristics.

3 MDM of Product Data

Even though this segment of MDM is generally called MDM of product data it involves much more than what would be usually understood as a product. This chapter focuses on features of this MDM domain, product data specifics and covers benefits of implementation.

This MDM domain is usually referred to as MDM of product data, another common term is Product Information Management (PIM). These terms are interchangeable and will both be used in this thesis.

For the purposes of this thesis the findings of the following studies will be used:

(1) Product Data Quality Report by TDWI
Source: [RUS2010] – not based on respondents
(2) Product Information Challenge by Dan Power
Source: [POW2009] – not based on respondents
(3) The State of Data Quality Today by The Information Difference
Source: [INF2009] – based on 193 respondents
(4) Trends in Product Information Management by Ventana Research[6]
Source: [VEN2012] – no data on number of respondents available
(5) The ROI Case for Enterprise study by Aberdeen Group[7]
Source: [ABE2011] – based on 634 respondents

Most organizations have several definitions for products, materials, parts, as can be seen in Figure 3. They also have multiple data elements and hierarchies that describe products, because they are often comprised of several parts and subassemblies. These can be also referred as products.

[6] Ventana Research is an audit and research firm
[7] Aberdeen Group is a research and consulting firm specializing in IT infrastructure

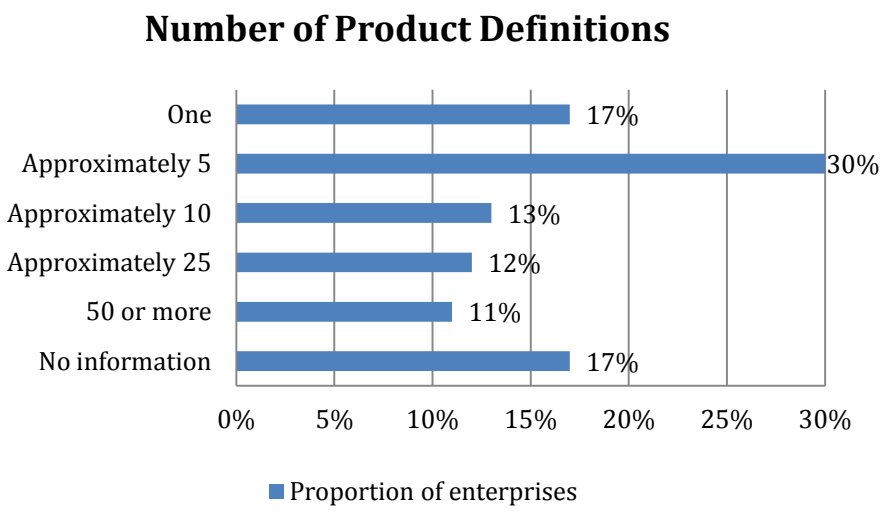


Figure 3 - Number of product definitions in enterprises [RUS2012] 8

For the purposes of this thesis the term product is used for a wide range of goods and services all across industries in both the private and public sector. It covers items from industrial devices, drugs and healthcare products, grocery items, electronics, building material, furniture, vehicles, financial service-based products, telecommunication services, assets, energy, social services, weapons, criminal justice, etc. The word product is also used for separate parts and substances of the above mentioned items, and raw materials.

3.1 Product Data Characteristics

In order to see the big picture product data and its unique domain needs, it is important to understand the specifics of the entity standing in the center of a MDM solution.

Pointing out main characteristics and differences between product and customer data, assists in the understanding of distinct needs in terms of MDM implementation, such as identifying processes, data quality approaches and standardization specific for PIM (product data cleansing, data model structure, etc.) as opposed to CDI.

[8] Based on 219 respondents with MDM experience

3.1.1 Variability

Customer data sets are much more static than product data. In case of product data, the structure is dependent on industry, product development processes, global and local context etc., while party data is usually of the same structure containing similar sets of attributes regardless of the industry.

This is also why the numbers of attributes are much higher when it comes to product data than in customer data. [POW2009]

3.1.2 External reference databases

There are many standard ways and resources for verifying person's attributes, such as official government statistics of names from registries (e.g. national standards for names or national postal standards) which can serve as a source for external reference data. This is not the case for product information. Generally product data is lacking standards as most manufacturers are reluctant to release details, because of their concerns of becoming commoditized on the internet. There has been progress towards standardizing some elements of product information in certain industries, with industry associations and government bodies promoting standards, for example United Nations Standard Products and Services Code (UNSPSC). [POW2009]

3.1.3 Categorization

It is more common for product data to require a hierarchical model than it is for customer data. This is because product reference data tends to be comprised of multiple sub-assemblies, and parts, which themselves require definitions. On the contrary, customer data works well with record-based modeling, although it may also be structured into various hierarchies, such as corporate family trees or sales geographic rollups.

Product hierarchies express sub-assemblies and units of the final product. Product hierarchy is often called bill of material (BOM)[9] and is therefore mostly used by manufacturers. An example of a product hierarchy is shown in the figure below.

[9] Bill of material (BOM) are the sum of parts and subassemblies which comprise the future product

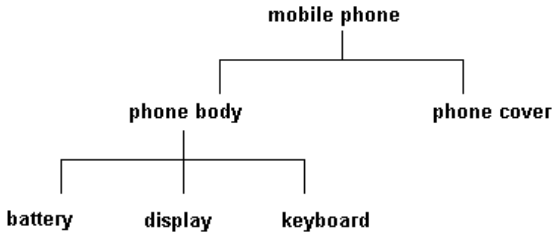


Figure 4 - Example of product hierarchy for mobile phones [AUTHOR]

Product categorizations tend to be quite complex, each enterprise usually creates its own model and therefore its own categories, there are no widely accepted standards. Generally, there is more than one product taxonomy (including product/product line etc.). [POW2010]

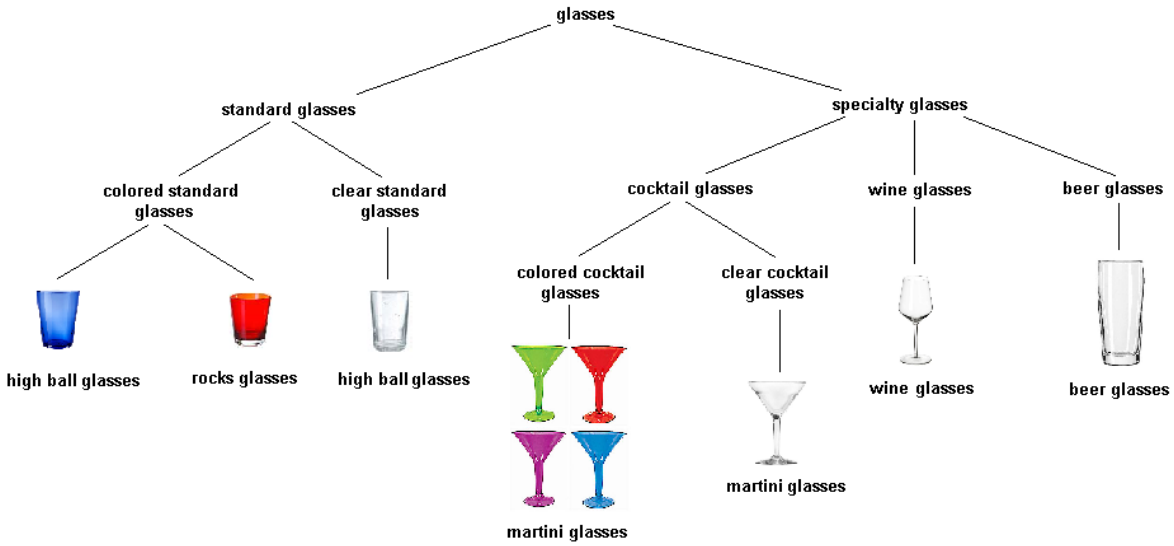


Figure 5 - Example of product categories for drinking glasses [AUTHOR]

There may be several product categorizations in a company due to inconstancies among its departments (each department using its own categorization) or as a consequence from company merges and acquisitions.

According to a study conducted by Ventana Research, 35% of organizations must manage 26-100 product categories and about 25% manage 101 to more than 1000 categories as is depicted in the following graph. [VEN2012]

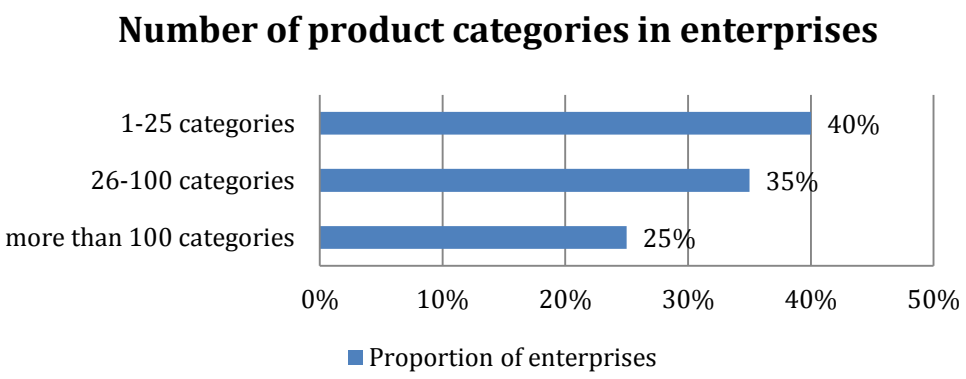


Figure 6 - Number of product categories in enterprises [VEN2012]

3.1.4 Structure

A lot of product data is unstructured (such as engineering or marketing documents and product catalogs), or poorly structured (like description fields overloaded with lots of information which is saved as separate fields like size, weight, color, etc.). This is not the case for customer data, as it is usually solely saved in company databases. [TDW2010], [POW2009]

3.1.5 Cleansing Variants

Data inconsistencies in product data are caused either within the enterprise or come from external sources:

- Bad data input on entry level within the enterprise
- Acquisitions and merges among organizations
- Expanding to global markets causing language and inconsistencies of measure
- Adding new suppliers to increase product diversity

The aim of improving data quality is to correct data from errors, and inaccuracies of order, to increase data reliability. It involves a set of processes with various data operations commonly called data cleansing, standardization, consolidation or enrichment. Nevertheless the whole technique is generally called data cleansing[10].

Data cleansing also requires a different approach for product and customer data. Since product can be identified as a vast range of product means from consumer packaged goods to electronics, it also

[10] Further information on data quality can be found in many dedicated sources. Perhaps the most know authority in this field is Larry English.

brings the variety of product attributes regarding their specification, technical information, product structures etc. Therefore while CDI requires an easier pattern-based approach, product data requires a semantic-based approach. This is because of its adaptability to various product categories, since each category of product data needs to be cleansed according to different rules for each type of product. [RUS2010]

According to a study from 2009 more than 50% of enterprises consider their data quality good or very good. This is quite surprising considering the fact that only about one third of the enterprises had a data quality initiative, see the figure below. [INF2009]

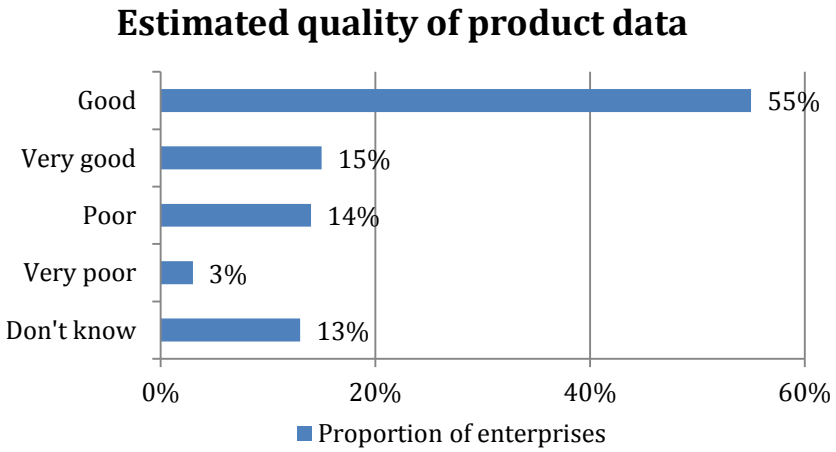


Figure 7 - Estimated quality of product data [INF2009]

Over 70% of enterprises believe product data it is difficult to standardize. This is not surprising given the complex nature of product data. A Figure 8 shows the detailed results of the survey. [INF2009]

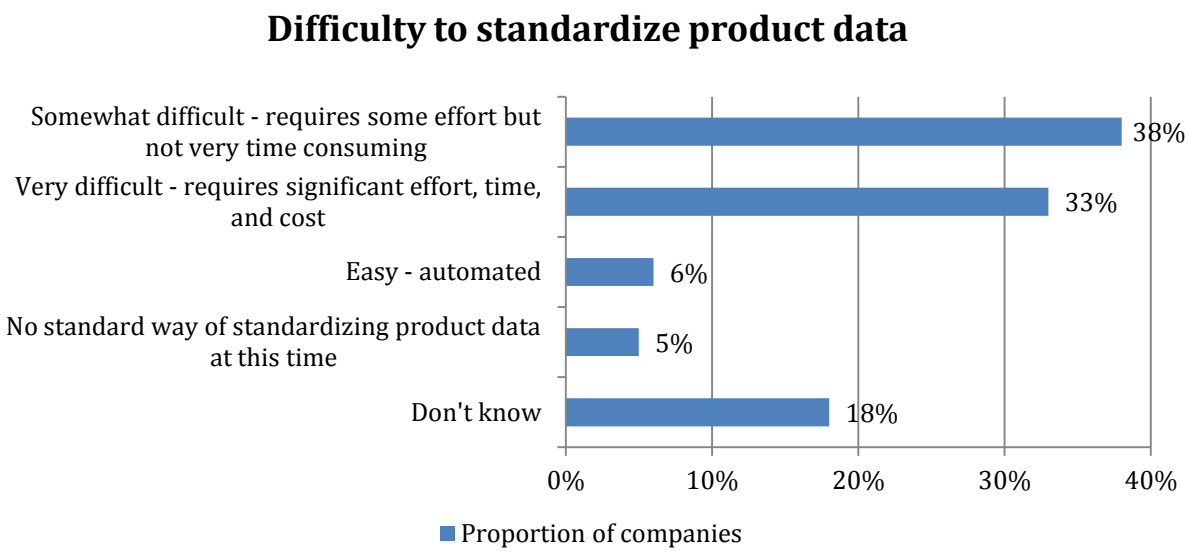


Figure 8 - Difficulty to standardize product data [INF2009]

Figure 7 in combination with Figure 8 clearly shows how much enterprises and their employees tend to be optimistic about the quality of their product data.

3.2 Defining PIM

Due to the complexity of a company's economic sector, Product Information Management (PIM) is a very wide field. The diversity of products and their specific needs for an industry's given attribute sets cause a differentiation in unique needs of various enterprises. This chapter focuses on PIM definition and briefly points out its basics.

Product data is stored and used in heterogeneous management information systems, to list the most common ones:

- Contract management
- Inventory management
- Accounting
- Invoicing
- Product entrance and termination

In order to use consistent and centralized product data, typically scattered across ERP systems, spreadsheets, PLM systems, or databases, organizations need to gather information from multiple

locations to a single centralized repository. Thanks to standardization, un-duplication and other functionalities of product data and its relationships, companies are able to efficiently manage, share and provide their product data through a stream of information from design, through manufacturing to product selling and servicing.

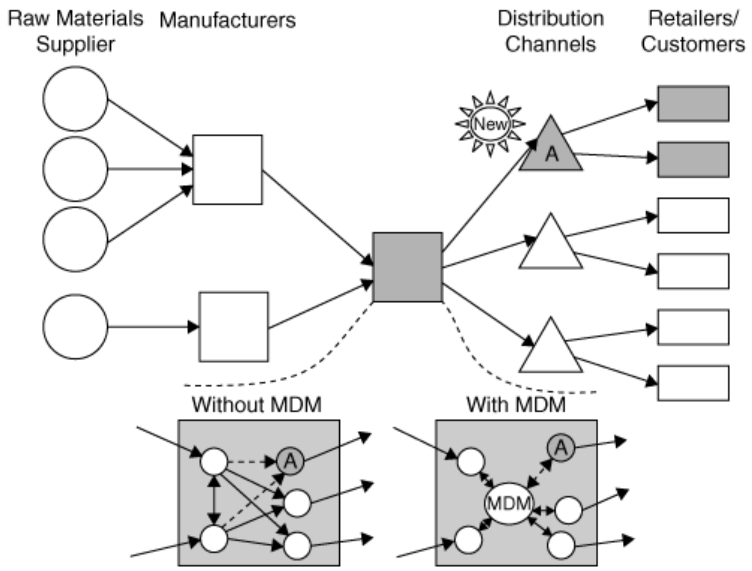


Figure 9 - Supply chain with and without MDM [DRE2008]

Figure above shows a view of a supply chain represented by a flow among its nodes from raw materials suppliers to manufacturers which transform raw materials into finished items. These may be then transformed into other items and distributed to wholesalers or retailers. At the very end are products bought by end consumers.

The benefits Product MDM data brings are demonstrated on a fictional enterprise being a node in the supply chain. This enterprise has two manufacturers and several distribution channels. In grey boxes below, the example demonstrates the difference between those enterprises which use MDM, and those which do not. MDM interconnects all its systems without the need to interconnect all systems among each other, but instead connecting all to the central repository.

Definition by Gartner:

"MDM of product data solutions are applications designed to create a single view of a product for an enterprise, across all operational and analytic uses, independent of any other repository of product data. MDM solutions store master data (and metadata, or both) related to products and other attributes of data pertaining to products. MDM systems can operate as a system of record (where

product and additional data is initially created and subscribed to by remote-consuming systems via messaging infrastructure), as a system of reference (where systems subscribe to remote master data using the MDM solution as a "look up" to locate and access the data) or as a mixed record/reference deployment." [GAR2011]

Product master data may include information about product ingredients, components, assemblies, usage, versions, patch fixes, pricing, discount terms, auxiliary products, manuals, design documents, forms and images and many others. Each enterprise has a unique set of product master data derived from its specific requirements.

3.3 Need of Product Information Management

This chapter describes Product Information Management in context of business needs and its use in organizations.

Organizations have their product data spread across numerous systems, applications and spreadsheets which are usually of different structures. According to a study by Ventana Research 57% of manufacturers and 38% of retailers use 4-10 different source systems to manage and store product information and 28% of all enterprises need to manage more than 10 different sources of product data. Therefore it is not surprising that more than 40% of respondents struggle with absence of a centralized product data repository, as well as too many incompatible tools and unsynchronized metadata stores. 78% of respondents are not confident in their product data quality and nearly 50% want to improve it. [VEN2012]

It is interesting to see how little attention is paid to product data quality and generally managing product data as it was ranked first in terms of priority from all data quality areas (in order product data, financial data, customer data) by respondents of The State of Data Quality Today study [INF2009].

Managing product data through a specialized MDM solution does not bring benefits to all types of organizations – one-fourth of organizations consider the business case for PIM insufficient for them [VEN2012]. This may happen, for example, in cases where a company has a small product portfolio, or due to generally smaller volumes of product data, since MDM is profitable only to mid-to-large sized enterprises with a greater amount of product data.

Enterprises seek PIM for multiple reasons. According to a study by Ventana Research most enterprises are mainly motivated by functionalities supporting product management and product life cycle, since product data is usually of very difficult structure as was already mentioned. More reasons of why enterprises implement MDM of product data can be seen in the figure below.

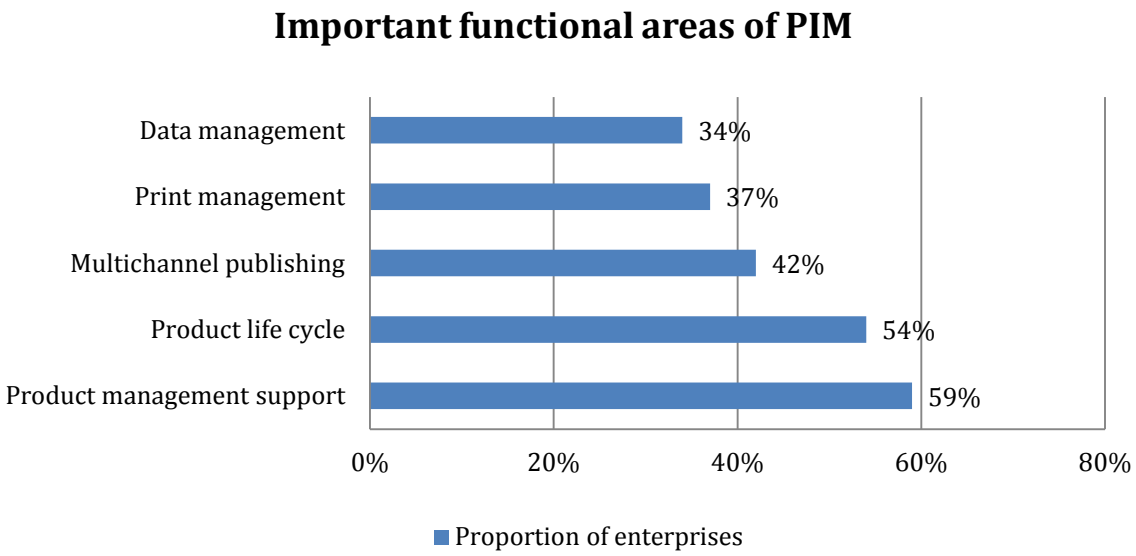


Figure 10 - Important functional areas of PIM solution [VEN2012a]

As mentioned in the previous chapter, some organizations develop home-brand PIM solutions themselves while other organizations adopt a dedicated solution. The overall share of dedicated software is only about one-third of all PIM solutions. According to a study by Ventana Research the share is 33% and according to a study by Aberdeen Group it is 27% of respondents. [VEN2012], [ABE2011]

3.4 PIM Implementation Benefits

The benefits of MDM have already been outlined in Chapter 2. In this chapter, MDM benefits are listed in the context of product data.

This chapter lists the most valued benefits as they have been found by respondents of two very recent studies focused on PIM. Further on, a complete list of benefits and their description can be found.

According to a study by Aberdeen Group [ABE2011] the most improved metric through PIM solution is the time spent searching for product related data, as can be seen in the table below.

Table 1 - PIM benefits [ABE2011]

Benefits	Improvement with PIM
Average number of hours per week an employee spends searching for product information	79% better
Year-over-year change in product development cycle	9% better
New product introduction that meets targets for time, volume and quality	6% better

According to a study by Ventana Research, preferences about the most valued benefits from MDM of product data vary. According to 47% of users, the most beneficial aspect is data quality improvements, while the second most popular contributions are related to better product data management (44% preferring potential of product cross-sell and up-sell, and 41% appreciate greater availability of product data across the enterprise) as can be seen in the graph below. [VEN2012]

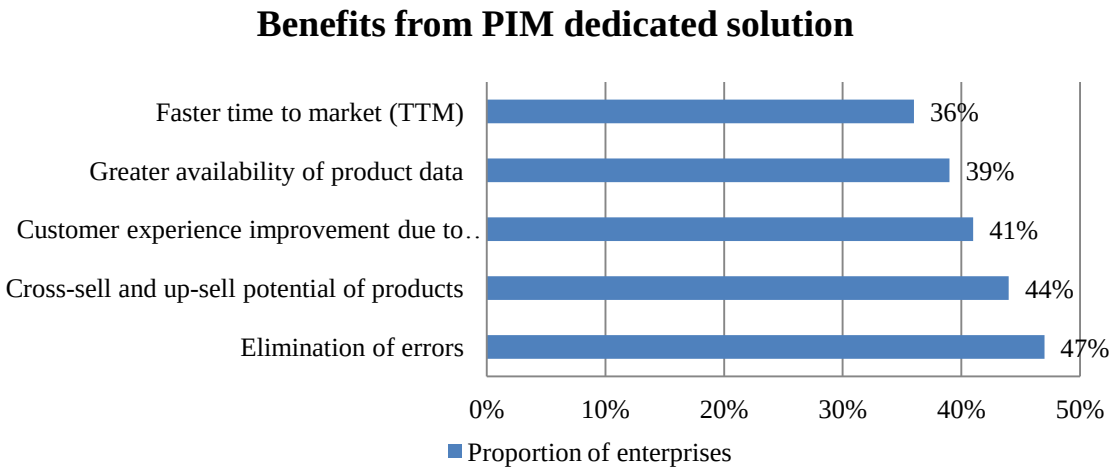


Figure 11 - Benefits from PIM dedicated solutions [VEN2012]

Findings in studies mentioned above clearly show significant impacts of PIM on processes within an enterprise and product data quality. The chapters below list and describe all benefits in greater detail.

3.4.1 System Interconnection

With a central master data repository, there is no need for point-to-point integration among enterprise systems, and applications. Instead all applications are connected to a central database. Since product data most commonly reside in various content management systems, databases, applications, PIM solution must handle different data formats at the input.

The graph below represents the most common formats of product data, according to a study made by Ventana Research. As was pointed out, product data is often stored in unstructured data formats. Using such data in PIM is a challenge in terms of tagging and data loading.

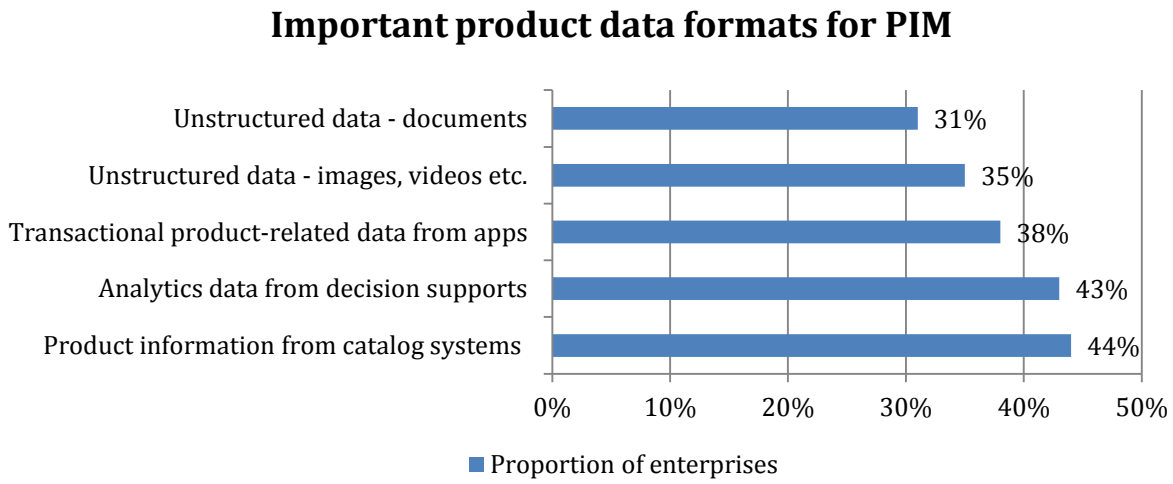


Figure 12 - Typical data formats for PIM [VEN2012a]

3.4.2 Multichannel Availability

Today's consumers require information to be accessible from several different channels at the same time, e.g. printed catalogs, retail stores, mobile applications, partners, dealer and vendor portals, smart cards, and/or websites. A recent study by IBM showed that 75% of customers prefer to first browse online, and then purchase the selected item in a store (did they give a specific region/country for this? Buyer preferences vary a ton depending on where you are). For 68% of customers, it is important to compare the price on a mobile phone as well. The same study shows that 51% of customers have purchased an item from a website of a physical shop. This shows the importance of sharing unified information across all channels, to follow increasing needs for multichannel communication and avoid customer dissatisfaction due to inconsistent data. [IBM2012]

Different channels have different requirements for channel specific information. This means that each channel might require a different attribute set of product data (e.g. an online catalog is going to require a different set than a dealer portal). There are also different requirements for a web page compared to a printed catalog in terms of synchronization. While it is important that a web site is automatically synchronized (either real-time or near real-time), printed catalogs only need to be synchronized before publishing.

3.4.3 Quality of Product Data

According to Dan Power[11], product data is of worse data quality than customer data on entry level with regard to ACT+C[12] assessment (accuracy, completeness, timeliness and consistency). [POW2009]

MDM solution provides product data for data quality processes, such as data assessment via data profiling tools to identify problems with quality of the actual data. Various data sources require specific approaches towards data cleansing and standardization. These involve data un-duplication, conversion, enrichment, translation (multi-lingual data support) etc.

In terms of product data conversion typical examples are:

- Weight conversion
- Currency conversion
- Speed conversion
- Size (in connection to packaging) conversion

There are several standard classifications for products providing single coding convention, which help enterprises deal with heterogeneous products and manage different product description styles. Identifying products, according to commonly used classification, helps to merge duplicate product records, share the same product identification with suppliers and business partners (only in case they use the same standard), and in some cases enrich product data with descriptions provided by classification standards or hierarchical classifications with several levels. Examples of commonly used international standards are:

[11] Dan Power is a founder and President of Hub Solution Designs, a global management and technology consulting firm specializing in MDM and working in this field over 20 years.
[12] There are several data quality classifications based on different aspects of data.

- eCl@ss – cross-industry product data standard for classification and clear description of products and services; mainly used in Europe; contains hierarchies [NET_04]
- UNSPSC (United Nations Standard Products and Services Code) – cross-industry product data standard for classification; mainly used in North America; contains hierarchies [NET_05]

There are numerous different classification standards for products, further information of the most commonly used ones can be found in the chapter Product Category.

3.4.4 Consistent Product Information across the Enterprise

Adding centralized unique product ID numbers across all departments is one of many aspects of centralized repository, which guarantees reducing incidents caused by inconsistent data. Typical examples are the adding of a product which is already part of the portfolio, or out of stock inconsistencies (41% of respondents are not satisfied with the ability to receive in-stock status while in the particular store, compared to almost 20% of respondents shopping online) or item tag problems (14% made no purchase because they could not find the desired item due to a badly placed item tag and 11% did not find the price tag. [MOT2011])

Centrally managed product data reduce the time, and costs, for producing and updating catalogs, online content, and other marketing materials by connecting product content with desktop publishing software, content management systems, and other programs.

3.4.5 Consistent Product Information with External Subjects

Providing trusted data used within the enterprise to its partners, retailers, suppliers and distributors helps to keep relationship information between products and their retailers and suppliers and allows for better control of supply chain processes. It also allows enterprises to trace historical product information which helps them to better understand their product lifecycle.

3.4.6 Consistent Reporting

Business decision making requires consistent product data distributed across all functional enterprise departments. PIM provides product data with error correction and a higher level insight into product data, which can be used for various analyses. Putting product analyses into context of other data helps them to identify up-sell or cross-sell opportunities for enterprise products, product sales per period, and others.

3.4.7 Ensuring Compliance with Regulations

There are many national and international agencies for the enforcement of food, drug, medical devices, cosmetics, radiation and other controls supervising the safety of people. Regulatory rules vary depending on the industry. Non-compliance with product related restrictions may cause penalties. Therefore validating enterprise product data against authorized values helps enterprises have an overview of their product contents and structure. Such regulations for different industries of product data are for example:

- Regulation on Registration, Evaluation, Authorization and Restriction of Chemicals (REACH)
- Restrictions of Hazardous Substances (RoHS)
- Regulation of medicinal products (EudraLex)
- Food additives law regulations

3.4.8 Better Search Results

There are several ways PIM improves better search results by both internal and external search platforms:

- Adding rich attributes and metadata to product data makes the data more likely to be found by search platforms and other tools.
- Consolidating products in case of duplicated items returns more corresponding results without duplicities.
- Cleansing product data attributes and classifying it under the correct groups or hierarchies.

3.4.9 Reducing Time to Market

In the current speed of new product introduction, it is important to bring a product to the market before competitors. PIM may speed up the process of launching new products by sharing a single view of consistent, high-quality product data inside and between enterprises and integrating product master data to all involved parties.

Figure 8 below represents a simplified scheme of new product introduction. Typical activities in the process are depicted with the responsible roles. The activities include product classification within the portfolio, product properties definition and assigning the product at all needed locations. It is critical for the PIM solution to reproduce the process in the same manner in order to comply with enterprise standards.

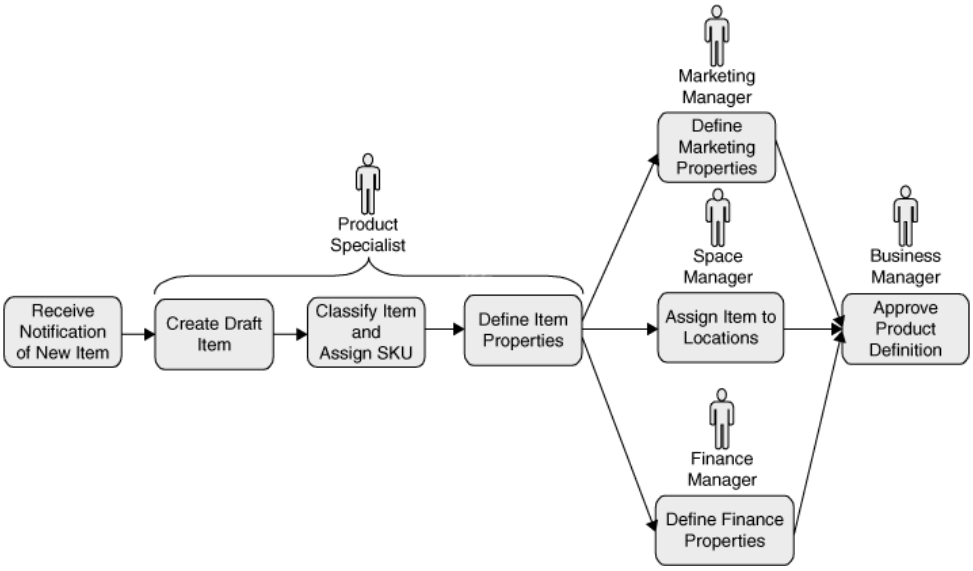


Figure 13 - New product introduction process [DRE2008]

3.5 PIM Market

As mentioned above the variability, lack of external standards, different requirements for classification and categorization and the differences in hierarchies are creating a great variability and distinction. This is the reason why most vendors doing PIM are specialized on certain industrial sectors or departments.

There are two main studies focused fundamentally on MDM of product data – Gartner's first analysis of the PIM market dates back to 2005 while Ventana Research launched its analysis in 2011.

3.5.1 According to Gartner

Gartner started to publish Magic Quadrant for both domains of MDM (CDI and PIM), but the differences in requirements and therefore software applications of either caused splitting the study into Magic Quadrant of Product Data and Magic Quadrant of Customer Data.

In the study each vendor is positioned into graphic representation of the PIM market based on customer feedback collected by Gartner. Ability to execute and the completeness of vision are the two key metric units, each represented on an axis.

Worldwide MDM software revenue will reach $1.9 billion in 2012 with 21% increase from 2011, with the project of MDM for product data to reach $688 million in 2012 and to surpass $1.1 billion in 2015. This trend shows that enterprises are aware of their distorted and poor quality data, often stored in separate silos, missing a central view of their data. [GAR2012]

According to [GAR2011] the market for product MDM is fairly healthy with $490 million with a 10% increase over 2009, which makes it actually bigger than the CDI (Customer Data Integration) market at $446 million. To compare, software spending on all MDM domains for the same year were estimated $1.4 billion with a 10 % increase. As seen below, enterprises in PIM market segment remained without major changes.

IBM, Oracle and SAP cover approximately 36% of the total MDM of product data software market spending for 2010. The remaining market is fairly fragmented. [GAR2012]

General trends in MDM are adopting multi-domain solutions and using software applications delivered by specialized vendors. [GAR2012]

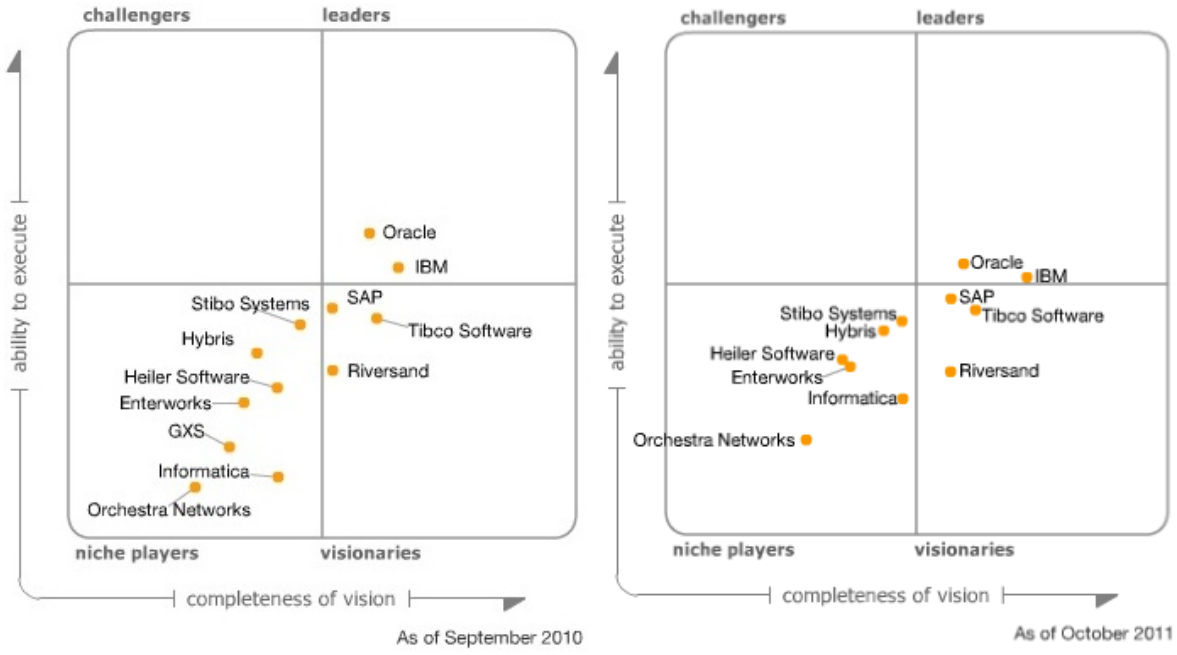


Figure 14 - Magic Quadrant for MDM of product data [GAR2010] [GAR2011]

3.5.2 According to Ventana Research

The Ventana Research Value Index is an evaluation of PIM software tools by vendors marking them with units from 0 to 100% from two points of view.

(1) Product related – includes 5 categories: usability, manageability, reliability, capability and adaptability. These categories are displayed on the Product axis.
(2) Customer assurance related – vendor validation category and TCO/ROI (total cost of ownership and return on investment) category. These are displayed on the Customer Assurance axis.

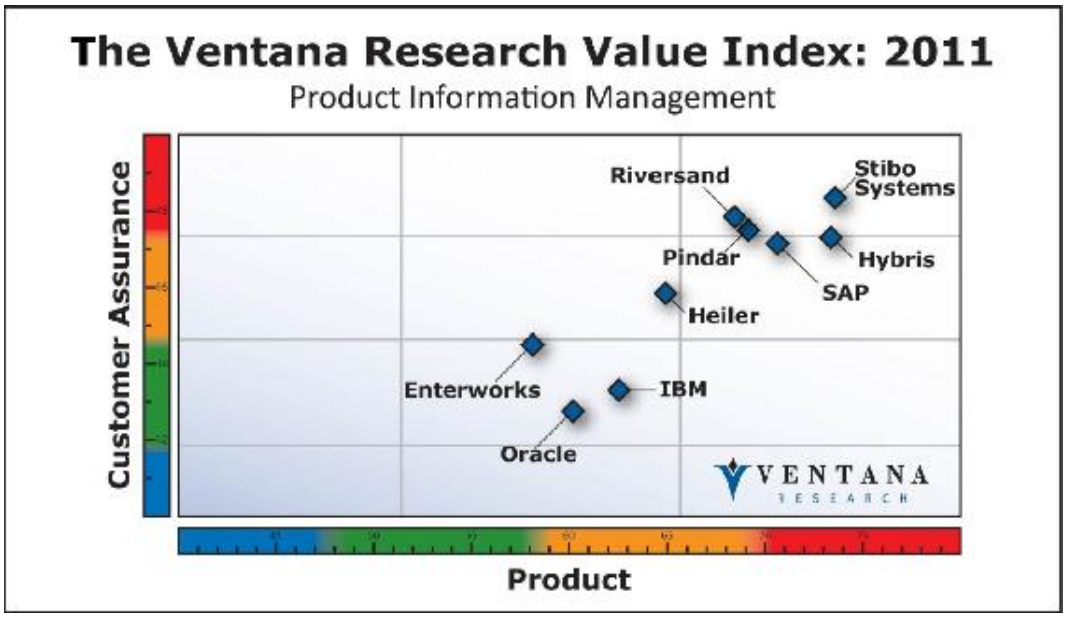


Figure 15 - Ventana Research Value Index 2011 [VEN2011]

Customer assurance related categories:

In 2011 depicted in Figure 15, where three vendors (Oracle, IBM and Enterworks) were placed on the scale of Customer Assurance between 12 and 15, and the remaining vendors were placed above 17 (except for Heiler). This has dramatically changed over year, as in 2012 all vendors (except for Oracle and IBM) were placed above 17 opposed to year 2011.

This shows an improvement from the customer point of view in terms of TCO and ROI assessment of most vendors.

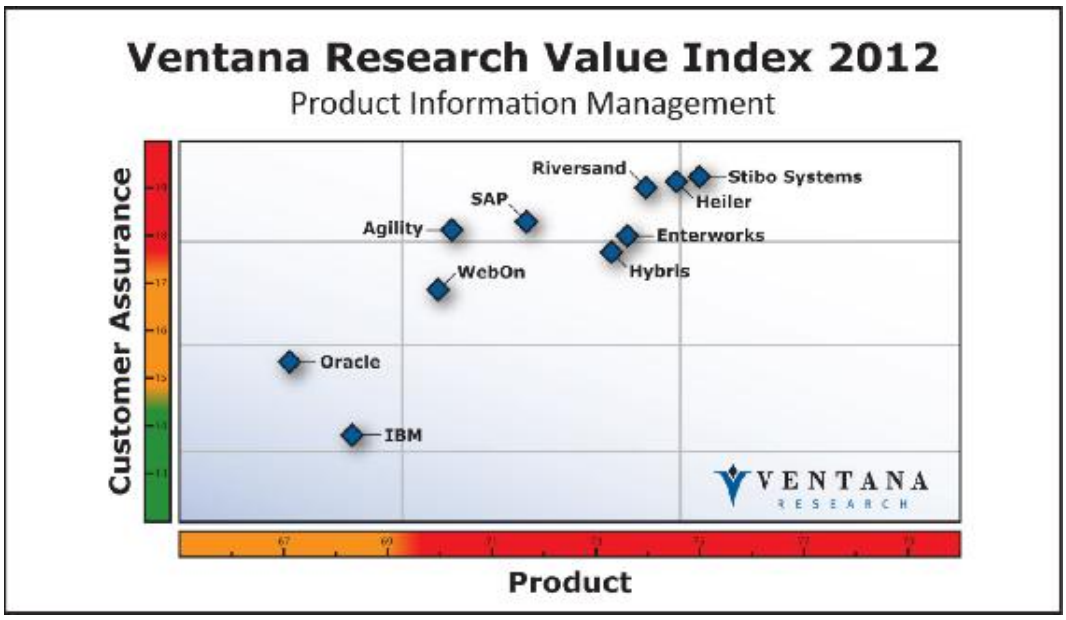


Figure 16 - Ventana Research Value Index 2012 [VEN2012]

Product related categories

Study by Ventana Research came to a conclusion that maturity of PIM solutions has increased as it is visible on the comparison of Ventana Research Value Index from both years. In 2011 all vendors were positioned on the product axis on scale between 58-73%, while in 2012 all of them were positioned between 66% and 75%. Stibo Systems and Hybris products were evaluated as the leading vendors in the product category in both years, while Heiler and EnterWorks have made a greater improvement. Oracle and IBM have been evaluated as vendors providing the least mature solutions both years. In 2011 EnterWorks solutions was evaluated as 68% and moved to 74% in 2012.

3.6 Chapter Summary

This chapter introduces typical features of product data, for this was chosen to use a method of comparing product and customer data from several perspectives in order to understand functionalities of product information management. Product data characteristics will be used in the practical part of this thesis.

Further on it contains a definition of Product Information Management (PIM) and also describes its main benefits regarding the main components of MDM in order to point out its contribution to the enterprise.

It also lists distinctive characteristics of most common MDM domains (customer and location) to lay out specifics of product data further in the text.

The rest of the chapter focuses on PIM market and current most notable vendors and trends. In order to give the most objective analysis the PIM market overview was based on main findings of two studies issued by different analytical firms presenting vendors with mature solutions.

4 Implementation

Each organization has unique data sources, infrastructure, sets of data and processes and therefore unique requirements for its PIM solution. This substantiates that every PIM solution needs to be tailor made for the specific enterprise.

Nevertheless some core aspects tend to take place in each PIM alias MDM adoption. This chapter aims at summarizing main areas based on findings in previous chapters and work experience of the author to create a basic methodology for PIM solution adoption. The methodology covers software tools, implementation processes in solution design and company background given by enterprise internal processes and is accordingly logically split into several parts.

It is not the aim of this chapter to describe architecture of a PIM solution as it always depends on the specific company infrastructure.

4.1 PIM Solution Components

As can be observed from previous chapters MDM of product data is a complex problem in most organizations and cannot be approached in a simple manner. PIM adoption comprises several crucial aspects, which were for the uses of this thesis divided into three main components:

(1) Best practices
(2) Application SW
(3) Enterprise background

A graphical representation of the methodology depicting all three parts is in Figure 17. On the bottom is the enterprise background representing company internal processes and environment, in the middle is application software representing the tools used for PIM and on the top is best practices reflecting the knowledge.

The purpose is to apply all best practices for implementation onto the specific software in context of previous distinctions (from chapter Enterprise Background) of an individual enterprise.

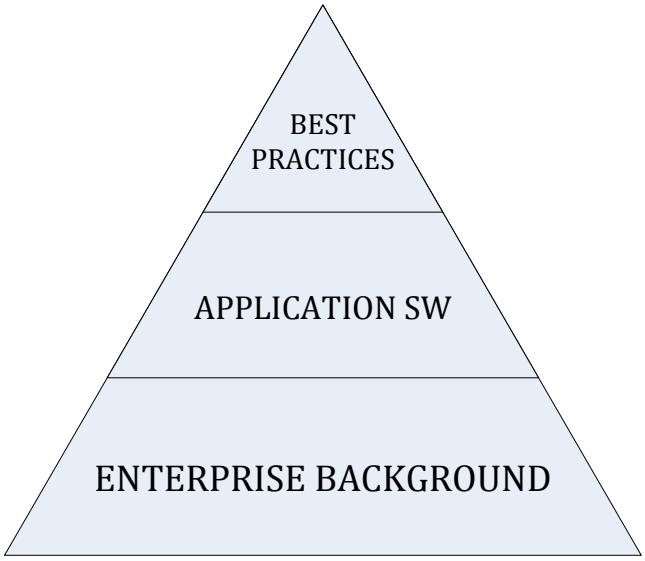


Figure 17 - Graphical representation of the methodology [AUTHOR]

The following chapters elaborate on the above mentioned components.

4.2 Enterprise Background

PIM is specific to enterprise processes. As each enterprise has a specific set of processes depending on its role in supply chain and the goods it sells, some patterns may be picked up.

There are several classifications of economic sectors and business subjects, but none of them are based on PIM benefits for certain types of industry. Regarding product hierarchy the situation is not better, as there is no internationally accepted hierarchy standard for products either. This chapter presents different types of classification standards and some examples of each type. It also summarizes the key aspects and benefits of each actor in the supply chain in order to differentiate various patterns in PIM adoption.

There are similar areas of focus depending on the field or industry of each enterprise. For the purposes of this thesis two kinds of enterprise group factors are used according to their influence on PIM implementation to narrow down the needs and benefits:

(1) Economic Sector – expresses business processes, which take place in order to manage product data
(2) Product Category – set of product attributes in relation to enterprise product industry

4.2.1 Economic Sector

Various actors in supply chain have different processes and needs for information depending on their role. Actors in supply chain cover everything from transforming raw materials into finished items, which may be then transformed into some other tools, distributing them from manufacturers to wholesalers or retailers and selling them to end users or consumers at the end.

Internal enterprise processes depend on products the enterprise is related to. It is the unique combination of the enterprise role in the supply chain connected with the specific features of the product which define the company processes (e.g. durable on non-durable products, such as food, relate to different product processes). Since all actors share their product data with actors across the supply chain, it is crucial that product data is reliable, uniform and of good quality.

In order to understand unique requirements for PIM of all actors taking part in supply chain, their main characteristics and needs are listed below in separate subsections with the exception of consumers (end-users). This can help enterprises classify main goals for their PIM adoption.

A Table 2 summarizing the requirements in greater detail can be found below the subsections describing supply chain actors.

Some enterprises may not comply with any of the classes below. This is given by the various business cases of each enterprise.

4.2.1.1 Manufacturer

Under this supply chain actor may be included raw materials manufacturers, part manufacturers and finished product manufacturers. Manufacturers are typically at the beginning of the supply chain and therefore each manufacturer typically uses its own standard ways to describe its product (unless it uses an internationally shared classification) both internally and in catalogs.

A manufacturer uses PIM for standardizing identification and description of its products in order to keep it consistent within inventory, shipping, sales, production management and material management and other processes. It is also essential to provide its customers with cleansed, standardized and enriched product data for better efficiency and reliability.

A typical aspect for a manufacturer is having an overview of what raw materials, parts are needed and are in inventory as well as understanding product hierarchy. This information typically goes

onto the product label together with other information, such as product name or description. This information needs to be standardized.

According to SBA (Small Business Administration's Office of Advocacy) [NEM2012] almost one third of all financial regulatory burden was paid by manufacturers (data relates to the US market only). The expenses are created by loads of paperwork, control and other insufficiencies. This number can be reduced by centrally managed regulatory managements and reports by implementing PIM.

Most commonly regulations are applied on healthcare, medical and drug products, chemicals, food and beverages, vehicles and electronics, toys, but mostly any product is bound with manufacturing regulations. Such regulation cover regulations related to the products themselves as well as environmental (e.g. pollution caused by manufacturing processes), workplace or tax-compliant regulations, standard product labeling and numerous others.

Typical representatives of manufacturers are Siemens, General Motors Corporation and Aero Vodochody.

4.2.1.2 Logistics provider

Logistics providers operate in all parts of supply chain among various actors focusing on distribution, warehouse storage, and transportation. This means that they must interact with many classifications and different data types. Another typical aspect of logistics is the need for accurate product as well as address data in order to deliver the right items to the right destination.

Another aspect typical for logistics is handling packaging sizes in order to comply with transportation rules and safety regulations as well as delivering and sending right sizes of packaged products. This means arranging products into the right volumes that fit the carriage and save space and money. Having an overview of products in stock without duplicities improves lowering inventory costs by not keeping exceeding numbers of products in their warehouses.

Typical representatives of logistics providers are DHL, Xpedex and FedEx.

4.2.1.3 Wholesaler

Wholesalers typically take over product descriptions and classifications from manufacturers and resell products to retailers. They also need to standardize addresses in order to track their products

as they may receive their address data from manufacturers from all over the world. Although it is not crucial to them to maintain customer analyses in high detail, they do need to pay close attention to in-stock status of products in order to avoid customer dissatisfaction caused by not having a wrong in-stock status. On the other hand it also helps lowering inventory costs by not storing big volumes of products. This can be achieved by consolidating duplicated product data.

Wholesalers usually sell products to industrial or business customers (typically retailers or manufacturers) in big volumes. This puts emphasis onto the performance characteristics of a PIM solution, which should be able to handle big amounts of data.

Typical representatives of wholesalers are Metro AG, Costco and Dacem.

4.2.1.4 Retailer

Retailers typically sell final products to the end-users. Therefore they need to handle numerous different product standards and classifications from wholesalers and manufacturers in order to have uniform product price tags or labels across the whole enterprise as well as accurate and reliable analytical reports (such as customer analyses, sales etc.) and accurate in-stock status in order to retain customer satisfaction. This requires transformation of multiple classification standards and value recognition in order to recognize the suitable data.

Depending on the size of a retailer, it may reinforce its own internal standards onto its suppliers (manufacturers and wholesalers) or share product classification standards among its partners.

Another very important aspect for retailers is communication with end-users via multiple channels, such as web portals, mobile applications etc. Therefore it is vital to keep all product data going through the channels synchronized.

Retailers also need to manage compliance regulations of product data. This is due to importing products from abroad since there may be different rules applied in countries of products' origin.

Typical representatives of retailers are Wal-Mart, Carrefour, Tesco and Amazon.

The table below describes all above mentioned actors in the supply chain and their specific requirements on the PIM solution in greater detail.

Table 2 - Enterprise PIM requirements based on its role in supply chain [AUTHOR]

Requirement	Manufacturer	Logistics provider	Wholesaler	Retailer
Synchronizing product data in subsidiaries	✓		✓	✓
Improving search functions by adding metadata			✓	✓
Using global categorization	✓	✓	✓	✓
Exchanging product information with external subjects	✓	✓	✓	✓
Correct labels on products (product description, subassemblies etc.)	✓			✓
Supplier data aggregation and consolidation	✓		✓	✓
Reducing Out-of-Stocks	✓		✓	✓
Converting pricing, sizing and other information	✓	✓	✓	✓
Consumer analytics (market basket etc.)				✓
Regulatory compliance	✓	✓	✓	✓
Classifying products for product catalog	✓		✓	✓
Packaging (regarding shipping)	✓	✓	✓	
On-shelf availability	✓		✓	✓
Hierarchy structures (BOM)	✓			
Multichannel availability				✓

Considering main requirements with relation with enterprise problems may help to identify the scope of implementing PIM.

4.2.2 Product Category

Product data is of diverse sets of attributes associated with different industries as opposed to party data which usually contains name, personal identifier, birth date etc. Data sets of product data depend solely on the industry and type of products with emphasis on different features by different types of data (descriptions, files, pictures). This means that for example furniture and medical products have sets of completely different attributes. This makes it difficult to find standards for product data opposed to address or party data, as they vary on each product. On top of this, sometimes competitors describe similar products differently in order to differentiate from other suppliers. Furthermore if enterprises have subsidiaries in several countries it is likely that they may be filled out in different languages.

As it was also mentioned in the chapter MDM of Product Data that product classifications vary from enterprise to enterprise depending on company portfolio, and sometimes there may be several classifications within the company based on departments or a combination of both internal and external classification.

Some enterprises cover only one specific type of product (for example specialty stores focused on selling only some types of products, such as food, cars, consumer electronics), while some need to cover products of various different types and needs (such as supermarkets covering industries from handicraft, through clothes).

All above mentioned problems turn standardizing product data into a real challenge because of no broadly accepted standard (recently an organization called GS1 has been trying to spread an international standard for product information by combining EAN and UCP [HUG2011], see Table 4). There are several product classifications based on different standards, but there is not a single official standard shared world-wide across all industries (although there are multiple international standards). Some classification standards cover products from one industry (e.g. pharmaceuticals) while other cover more industries or all industries across the supply chain. Neither detailed description nor evaluation of all available standards is among the aims of this thesis.

In relation to product classification, there are two kinds of approaches regardless of the size of product portfolio an enterprise can choose from:

(1) Using a classification standard by external institution
(2) Using an internal classification created by the enterprise itself

The table below lists advantages and disadvantages of both approaches.

Table 3 - Advantages and disadvantages of using internal and external product categorization [AUTHOR]

	Advantages	**Disadvantages**
(1)	Taking over a functioning standard	Choosing among all available standards (some standards are supported on country level, some on international level, etc.)
	Sharing unique product IDs with other actors in supply chain (under the condition that they use the same standard)	Being dependent on the structure of external classification, also in case of internal company changes as well as changes within the classification (different classifications have different levels of detail)
	Enriching enterprise product data with extra values provided by external institution	Financial aspect (some standards require license fees)
	Level of detail as some classifications focus only on a given segment and are very specifically oriented (for example high tech) while others cover all industries	Synchronizing classification versions (many classifications are of many versions due to upgrades), while for some classifications versions become obsolete and for other classifications several versions may be valid at a time)
	Advantages	**Disadvantages**
(2)	Level of detail is tailor made and therefore suitable to the given enterprise	Creating a self-sufficient classification is a very complex task in context of possible future portfolio changes
	No dependence on external structures and changes	It is not possible to enrich product data from external sources
	No payments to external companies	No sharing product IDs with other subjects in supply chain

Before making the final decision about adopting a classification standard by an external institution, enterprises should closely analyze the data quality of the external data source. It would be wrong to assume that all standards are of good data quality. Examples of classification standards in Appendix 10.6 present some data quality issues (see the example of Harmonized System).

Table below gives an overview of the most common international cross-industry product classifications and their structure. There are various categorization standards, most of them split

into several subtypes (some have for example an international version as well as a localized version).

Many changes in the field of international category standards can be expected in the next years as they will be probably adopted by more manufacturers as well as spread to more countries.

Table 4 – Examples of cross-industry product category standards [AUTHOR]

Standard Name	**Institution**	**Structure**
Central Product Classification (CPC) [NET_08]	United Nations	5 levels
Industry Classification Benchmark (ICB) [NET_12]	Financial Times and Stock Exchange Group	4 levels
Global Product Classification (GPC) [NET_09]	GS1	3 levels
Kompass Classification [NET_10]	Kompass International	3 levels
Harmonized System (HS) [NET_03][13]	World Customs Organization	2 levels
eCl@ss [NET_04]	The eCl@ss association	4 levels

Apart from category standards there are also several institutions creating an international database of unique product identifiers. In some cases there is a very thin line between these two types.

Most widely spread is GTIN, which works as a certain umbrella for international product recognition. It involves both UPC and EAN schemes, which were combined, and from 2005 on are used in combined format. This format has a form of a string of typically 14 digits (UPC) in North America and a string of 13 digits (EAN) + 1 check digit in Europe. [HUG2011]

Other formats of GTIN are also widely accepted. GTIN may be a string of 8, 12, 13 or 14 digits, all of them with the equal meaning but of a different encoding. Both EAN and UPC can occur in various variants as well (for example UPC-B, UPC-D, EAN-8). [NET_05]

[13] Harmonized System is used for international trade. Over 200 countries have adopted HS. [NET_03]

Several examples of industry specific standards can be found in Table 6 (both category and unique identifiers). It also includes the institution, which is responsible for the given standard and its basic structure:

Table 5 - Examples of industry specific standards [AUTHOR]

Industry	Standard Name	Institution	Structure
Pharmaceutical	Anatomical Therapeutic Chemical Classification System (ATC)	World Health Organization (WHOCC)	5 levels
	National Drug Code (NDC)	Food and Drug Administration (FDA)	3 levels
Media	International Standard Serial Number (ISSN) [NET_01]	International Organization for Standardization (ISO)	String of digits
	International Standard Book Number (ISBN) [NET_13]	International Organization for Standardization (ISO)	String of digits
Automotive	IMDS (International Material Data System) [NET_07]	Consortium of enterprises	String of digits

4.2.2.1 Regulatory Compliance

Based on product industry and/or category, a set of regulatory compliance rules may be reinforced on enterprise products. Some basic examples were mentioned in the chapter MDM of Product Data.

There are many compliance standards both international and national. There are compliance standards applied on all industries or just specific industry sectors, specific products or even specific parts of products, such as labeling, bar codes or packaging. Most common standards are issued by ISO (International Organization for Standardization). Some examples of ISO standards related to products are:

- ISO 9001, ISO 9002 and ISO 9003 – related to all industries
- ISO 13485 – specific to medical industry
- ISO 16949 – specific to automotive industry

An example of a local standard reinforced in EU is REACH (Registration, Evaluation, Authorization and Restriction of Chemicals) related to chemical products.

The above mentioned examples are to provide an idea of various regulations enterprises need to comply with and which could be therefore handled by PIM. The list of specific regulations depends on each enterprise, its processes and industry.

Many regulations are country or area specific, therefore when importing products or subassemblies from abroad each supply chain actor needs to assure, that the products comply with the country regulations. There are some ways ensure and verify, what rules apply on which products in order to protect its consumers after importing. An example of such TARIC code (Integrated Tariff of the European Communities), which is a 10 digit code designed by the European Union. It represents which rules apply on the given products when importing them to EU. [NET_06]

4.3 Application SW

In chapter MDM of Product Data is mentioned that about one third of all PIM solutions use a dedicated software tool by an external vendor. Nevertheless the trend in PIM and in MDM generally is to move from homegrown solutions to ones delivered by a specialized vendor with the necessary know how.

Most significant companies offering PIM application software on the global market were outlined in the chapter focused on PIM market. A more thorough list of vendors can be found in appendix 10.1.

PIM application software is a fundamental part of each MDM solution with defined functionalities. If best practices are applied in context of enterprise processes, together they form a successful implementation.

This chapter covers a list of core software tools for general MDM solution with regards to specifics for product domain. Mature PIM application software should contain the following functionalities regardless if they are covered by one or several different software applications. The list below elaborates on functions of each.

4.3.1 Data Model

In the center of each PIM solution stands the underlying master data model, which represents its data entities, their relationships and data attributes. These express a product definition across all departments of the enterprise. There may be one or several models of the solution present.

Each solution model typically comprises the following types of tables:

(1) Master data tables (e.g. master product table) – contain consolidated data set of master data related to product. In each PIM solution the master product entity stands in the center and depending on customer requirements other tables (such as supplier data) may be mastered. An example of a product master table can be found in appendix 10.2.
(2) Shared reference data tables (e.g. supplier address table) – each data model consists of a canonical representation of company data regardless of its structure in the source system.
(3) Internal reference data tables (e.g. product status table, product type table) – each company has its internal ratings or values for specific attributes. Internal reference data tables can be found on master level. They contain values only related to master tables. An example of such tables can be found in appendix 10.3 and 10.4.
(4) External reference data tables (e.g. outside product classification standard) – they are provided by an external subject, which establishes its structure.

To build a comprehensive model for PIM, the modeling environment must support functionalities suitable for modeling product data with all its mentioned aspects:

- Hierarchy management in order to represent complex product classifications and hierarchies which allows moving product across categories and hierarchies.
- Large comprehensive attribute sets as product data can contain hundreds of attributes. They support the extensive volume of product features reflecting different data structures in context of product industry.
- Support of various data types. This is because product data often contains product images (CAD files, images, etc.) or documents with unstructured product data.
- Modeling complex relationships among modeled entities.
- Data models can be automatically generated.

Below is an example of a hypothetical conceptual data model of a wholesale business. It shows a product entity with a basic product description and its relation to pricing, product categorization and orders.

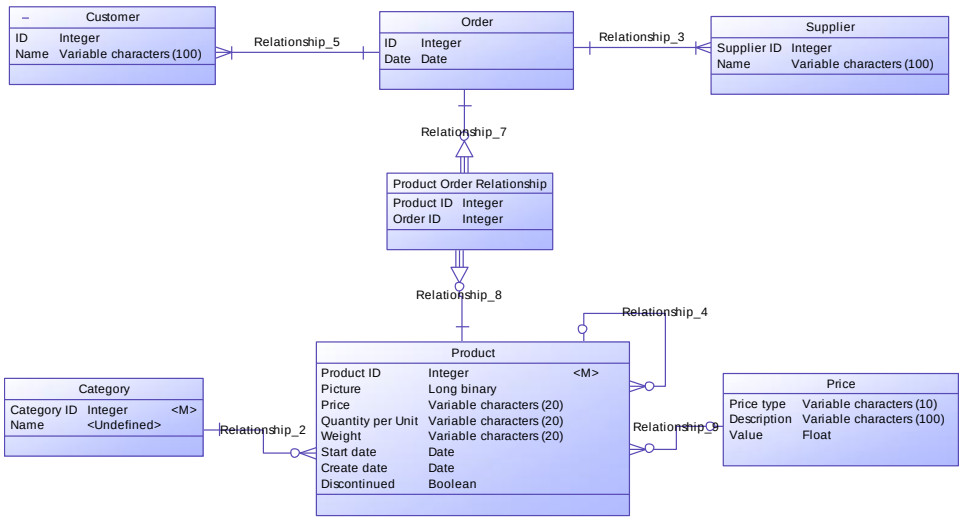


Figure 18 - Example of a conceptual model of reference data [AUTHOR]

4.3.1.1 Metadata Model

As product data is very complex, it is common to store in the model attributes in the form of data structures, containing more flexible information. These structures are then stored and decomposed on demand. This avoids replication of information and huge sets of records, which would slow down performance of the solution and waste resources.

Metadata modeling is especially important when working with hierarchies as it decreases complexity of a structure. The metadata model needs to be expansive and easily adjustable. Nowadays one of the trends in MDM is that solutions are moving towards being metadata driven. This means that a solution may be generated on the base of metadata. In such cases, it is the Meta API which operates over the data. The Meta API would provide information on the queries and data fields available for the entities in the PIM repository.

The tool should strive for easily interacting with the metadata. There should be ways to access the metadata and the information that is conveyed by it (such as the Meta API) to operate on the PIM repositories.

4.3.2 Workflow Management

Workflow orchestrates steps, tasks and operations for maintaining data within the solution automatically by placing them in order. This allows the solution to react flexibly.

Different workflows carry out processes related to various areas. Typical examples can be:

- Data loading
- Data quality processes
- Providing data to its consumers

Workflows are typically driven by a scheduler or by an event. A scheduler is a tool which manages their order and timing. In this context if an event starts a workflow, it is started manually (e.g. if a file is delivered, a workflow is automatically started to load the data into the solution). Depending on the functionality, workflows may run in parallel.

In order to adjust the PIM solution to enterprise requirements PIM solution must allow creating a custom workflow and building complex connections among many separate workflows.

An example of a graphical interpretation of a workflow is in the figure below:

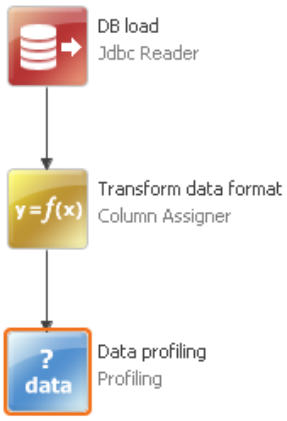


Figure 19 - Example of a simple workflow [AUTHOR]

4.3.3 Data Quality Management

As was stated in previous chapters, the core of each MDM alias PIM solution is made up by data operations such as data cleansing, enriching, validating, scoring, and consolidating. Therefore each MDM must have a sophisticated DQ tool.

It is quite common for enterprises to do some basic data quality monitoring and controls. Some enterprises even do more or less sophisticated product data cleansing. Data cleansing is a discipline on its own meaning that enterprises may have a complex data quality management implemented into their data processes having a standalone function without any MDM solution. In order for each MDM solution to have a significant impact on enterprise data, the core function must be data quality processes.

The function of a data quality tool is running complex analyses on data as well as processing of data, which is validated or corrected based on sets of rules or by comparing values with internal or external standards.

Generally, it is essential for each data quality tool to manage the following core functions:

(1) Complex data analysis for assessment, monitoring and reporting data quality through statistics. Typical examples of data analyses are:
   - Frequency analysis
   - Pattern analysis
   - Domain analysis
   - Analysis of minimum and maximum values

(2) Cleansing functions for verifying, correcting and enriching data, such as:
   - Verification against external reference sources
   - Enrichment from external reference sources
   - Data type recognition and validation (date x string)
   - Data scoring – for identifying data quality of a specific value
   - User defined and configurable validation rules
   - Semantic-based value recognition to enable accurate parsing and standardization

(3) Matching functions – the goal of matching functions is to identify and consolidate two instance records into a single master record.
   - Semantic-based value recognition

- Selection process for assigning duplicate records
- Data quality assessment in order to choose the best option from among all values
- Enriching from several duplicate records (for example if two duplicate records have a few vacant fields, the resulting unique record will blend all available values)

Most important technology requirements from the view of product data for data quality software are:

- Handling large attribute sets as products tend to be described by many features
- Handling big volumes of data as product data tends to be of much greater volume than customer data
- Handling big sets of specific cleansing rules for all attributes and mainly due to variable product categories as each category has unique requirements for attributes (mobile phones and printers are defined through different sets attributes)
- Looking up data in large external reference databases in terms of performance such as product databases (classification standards and regulatory compliance) because of big volumes of such standards (some of them containing hundreds of thousands of records)
- Understanding complex classification standards as well as converting among them
- Excellent semantic functions (e.g. parsing) due to semantic approach towards product data cleansing. The other use case is amount of products per package, as various packaging sizes contain different number of product items or volume (for example a bottle of water can be of 0,5l, 1l, 1,5l etc.). Each product category uses unique abbreviations, terms, vocabulary, sizes etc.
- Conversion – converting from different currencies, various systems of measurements (metric system, US customary units) in terms of product volumes, package content or packaging sizes (for example a bottle of water can be either of 0,25l or 250ml).
- Translating based on dictionaries – as companies expanding abroad often have problems with values in multiple languages
- Handling unstructured product data through metadata management
- Auto-learning to handle the product data in terms of parsing and standardization. As product description is reflected in product attributes, they may be of infinite values. Therefore some basic functions to recognize or read values and understand them would save time creating new cleansing rules.

4.3.4 Data Integration

Product data is present in many processes and locations (databases, web sites, catalogs, and many other) in an enterprise. Therefore it is typically stored in various formats.

PIM aggregates and loads desired product related data into the solution as well as it propagates standardized and consolidated data from the solution into applications and different channels. Therefore it should support importing from structured and unstructured data by adding metadata. Typically used data formats of product data are for example:

- Text formats .xml, .xls, .csv, .txt
- Database structures
- Binary files for graphic formats (as product data tend to contain product pictures or CAD drawings) or documents and forms
- Metadata of both structured and unstructured data

The software tool should also support different types of line breaks and file encodings (e.g. UTF-8, Windows-1250, ISO-8859-2).

4.3.5 User Interfaces

As mentioned before, MDM is an ongoing initiative which does not stop with the solution implementation. Therefore it should provide users (data stewards[14]) with an intuitive graphical environment so that they would be able to easily build, change and maintain the solution, such as:

- Analyzing data quality
- Adding and changing products
- Adding, changing or removing cleansing/matching rules
- Managing data quality findings (some data changes may require manual approval)
- Adding, changing or removing workflows
- Changing the solution model

[14] Data steward is a general name for a role, which is responsible for long term maintenance of the solution, such as altering solution model, cleansing and matching rules, managing ambiguous records and other related functions.

4.3.6 Security

Accessing the solution by various users could cause security issues; therefore the solution should offer some kind of user management for role or group-based access. This would grant “read” or “write” permissions to change the solution to only a selected group of users or possibly could even restrict some areas of the solution content depending on the user profile. Yet still the solution should provide keeping track of all changes. This should be maintained by a reliable version control tool.

4.3.7 Version Control

Version control systems allow keeping track of development of a PIM solution in a repository that it is integrated with. This for example involves adding new attributes or changing cleansing rules. Every change (revision) is recorded and archived with date, author and the differences towards the previous version.

Commonly used version control tools are e.g. SVN, CVS and GIT.

4.4 Best Practices

This chapter engages into describing a process which captures the activities connected to PIM solution definition and design process. Some activities are identical in any other data domain. Nevertheless all activities are described in context of product data needs.

Best practices were taken from previous projects focused on Customer Data Integration and adjusted to product data specifics derived from previous chapters.

The PIM solution design process can be customized based on specific enterprise requirements which form the scope of each projects. The process of designing a PIM solution should be followed closely as the order of all activities is usually based on outcomes from the previous activities.

This chapter does not take into account assigning roles for the below listed activities neither solution testing as it must be automatically involved into any software project involving software development. Therefore it is not necessary to point it out separately.

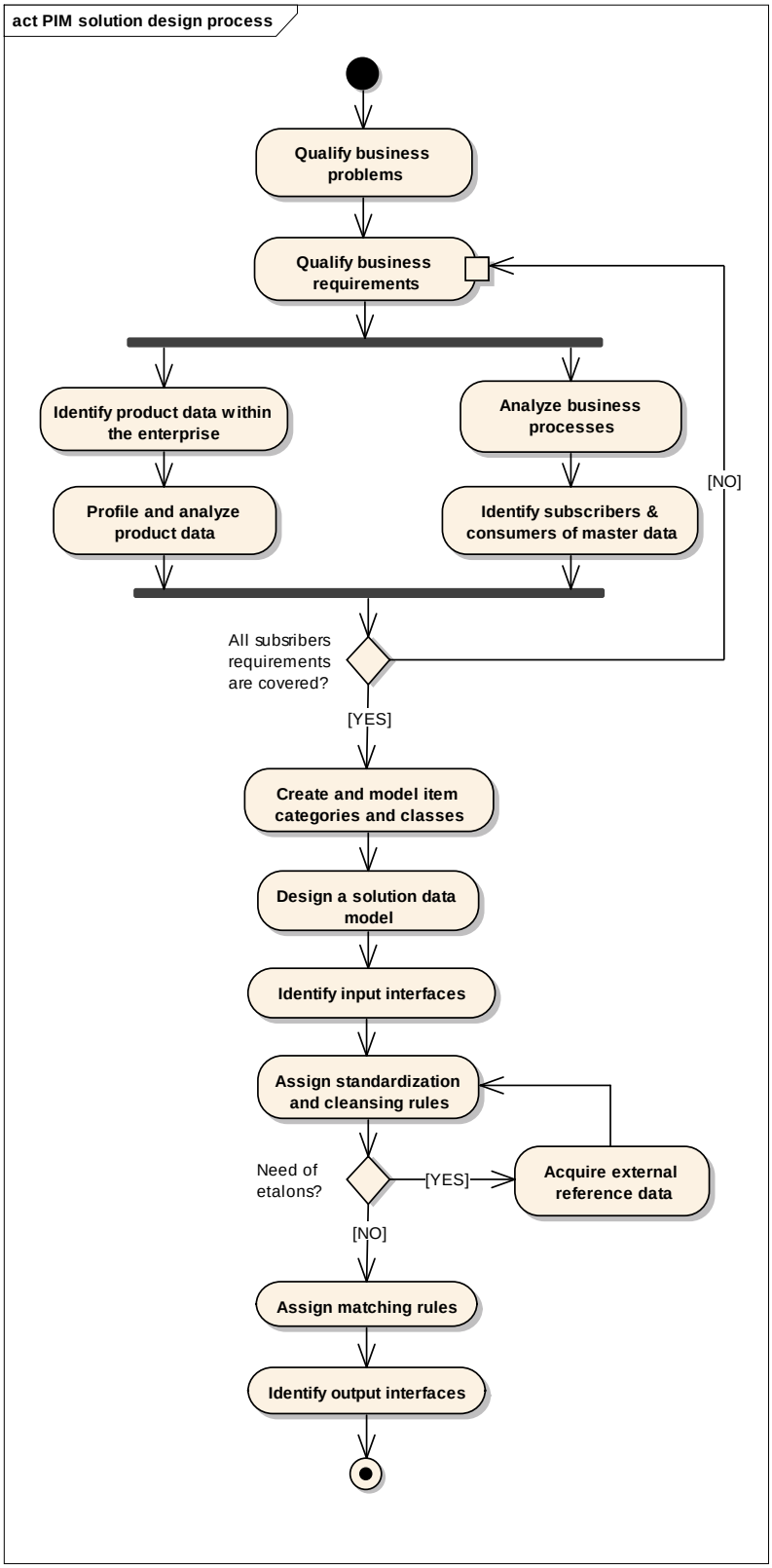


Figure 20 – PIM solution design process [AUTHOR]

MDM is an ongoing initiative, but implementation is a project. After finishing the PIM implementation, processes to adjust the PIM solution to new future requirements (such as adding new attributes or product categories) should continue, but are not covered by this thesis. The process representing PIM solution design is split into numerous activities, which are further described in individual subsections below. Activities listed and described in the next sections should be followed according to the succession in the figure above.

4.4.1 Qualify Business Problems

As mentioned in previous chapters not all enterprises have a use case suitable for MDM of product data. To figure out whether benefits from PIM adoption would outweigh the expenses, enterprises need to focus on qualifying and calculating their problems which may be managed by PIM (see chapters PIM Implementation Benefits and Economic Sector). Main business problems are usually connected to poor data and inconsistent processes. To mention a few examples:

- Wrong volumes of products in stock (e.g. sales of discontinued products, products not being in stock, sales of products of two or more different names, price mismatch, these all cause customer dissatisfaction)
- Too much work spent on synchronizing several sources to produce catalogs and communicate with suppliers
- Discrepancies in data analytics (duplicate product items)

Outcomes from this activity serve as a foundation for the follow-up activity:

(1) Qualify business requirements

4.4.2 Qualify Business Requirements

It is important to make a list of requirements derived from business problems, analyze which business processes are most affected by poor data, set priorities on which of them have highest potential to improve and set main challenges in order to create a sustainable long term PIM program in terms of scope.

Outcomes from this activity serve as a foundation for the follow-up activities:

(1) Identify product data within the enterprise
(2) Analyze business processes

4.4.3 Identify Product Data within the Enterprise

Depending on scope of the project, in this activity enterprises must localize which systems contain product data, their context and usage, metadata, images and other content associated with product data (both structured and unstructured) and its consumers and channels.

Within this activity enterprises also need to localize internal reference data related to its product data and possibly other data apart from product data derived from previous activities, which need to be analyzed (e.g. product vendors, list of countries of product origin).

Outcomes from this activity serve as a foundation for the follow-up activity:

(1) Profile and analyze data

4.4.4 Profile and Analyze Data

The term "data quality" is broadly used, but understood differently depending on the author. Generally data is viewed from several aspects, such as: accuracy, consistency, completeness, timeliness, currency, conformance and referential integrity. This thesis doesn't focus on all aspects on data quality, instead lays down approach to understanding product data.

Some enterprises already have data quality processes (from basic to more advances data quality handling) implemented into their company systems. In order to understand the present data quality problems it is necessary to run various analyses on the located data, set data quality metrics and start an ongoing reporting to observe data quality improvements.

Firstly it is important to define in which processed are which product data created and used, relate their metadata and internal reference data. Afterwards the data must be profiled using an appropriate DQ tool with the option running various statistics on the data. Based on the statistics data are evaluated. Commonly the outcomes of such profiling activities are statistics of:

- Null (empty) values
- Data types (boolean expressions, dates, strings)
- Dummy values
- Highest and lowest values
- Referential integrity among attributes, entities, systems

It is crucial to understand the localized data before any further steps are taken.

Outcomes from this activity serve as a foundation for the follow-up activities:

(1) Model Item Categories and Classes
(2) Design Solution Data Model
(3) Identify Input Interface
(4) Assign Standardization and Cleansing Rules
(5) Assign Matching Rules
(6) Identify Output Interfaces

4.4.5 Analyze Business Processes

According to Dan Power *"from an information management perspective, the process of developing, producing, distributing and supporting products is tremendously complicated when there's no one place you can go to see the entire product catalog for the whole enterprise."* [POW2010]

Nevertheless understanding the product lifecycle, processes connected to materials, parts, products, and other data used in enterprise processes and systems in an enterprise is vital for creating a data model for the future PIM solution.

Enterprise processes vary depending on many factors, for example durable on non-durable products have different product lifecycle, or seasonal clothes have different lifecycle than drinking glasses. Another important aspect is which supply chain actor is the specific enterprise.

Outcomes from this activity serve as a foundation for the follow-up activities:

(1) Model item categories and classes
(2) Design solution data model
(3) Identify input interface
(4) Identify output interfaces

4.4.6 Identify Consumers of Master Data

Firstly the enterprise needs to locate which systems will use the product master data, how it will be used as well as their unique needs for data sets. This is important, because based on the requirements of all subscribers, the source data will be chosen and managed. Another key aspect is

the frequency of publishing the data for the consumers, which will influence the frequency of updating the master data repository.

There are two types of consumers depending in which role they have in relation to the enterprise, examples can be found below:

- Internal – CRM, BI, marketing department, web-sites, catalogs, in-store operations etc.
- External – portals of partners (vendors, suppliers), audit firms, government, etc.

Outcomes from this activity serve as a foundation for the follow-up activity:

(1) Identify Output Interfaces

4.4.7 Model Item Categories and Hierarchies

As was already mentioned, each enterprise usually has its own internal product classification. In some cases there may be more classifications present or a combination of both internal and external categorization. The goal is to unify all present categorizations in order to keep consistent data.

The enterprise needs to choose the most suitable categorization standard based on:

- Enterprise industry
- Standards used by business partners

The main decision an enterprise needs to take is selecting whether it will be using the current or a new (either internal or external) categorization.

The decision between an internal or external categorization is based on comparing their advantages and disadvantages listed on chapter Product Category as each enterprise has unique requirements (e.g. business partners use an international standard, specialization on a very specific product).

When creating an internal categorization (either because the enterprise unifies several former internal ones or because the former categorization was found inadequate) it is usually crucial to consider the enterprise portfolio. In case of adopting an international categorization it is important to analyze if any external business partners use which categorizations and analyze selected external categorization. It has been mentioned that even though they may be standardized, there is

a chance of questionable data quality of the external categorization., which would in the end make it more complicated to keep consistent data creating additional workload to standardize and clean the categorization.

After selecting the categorization for the PIM solution, it must be modeled according to its internal structures.

Outcomes from this activity serve as a foundation for the follow-up activity:

(1) Create Solution Data Model

4.4.8 Create Solution Data Model

After analyzing product and possibly other data (activity Profile and Analyze Data), which were selected to be loaded into the solution, it is important to state which kind of models will be designed and used. Each model depends on the type of organization, its processes and product data (based on activity Analyze Business Processes).

A data model is usually iteratively built up by adding its elements during the implementation process depending on the requirements of the enterprise and master data consumers. This is given by the fact, that requirements tend to change over time.

Entities, their attributes and relationships can be designed after establishing the desiredmodel types. When choosing which entities and attributes will be modeled, it is recommended to keep the model simple because of big complexity issues related to performance (product attribute sets tend to be large). It is derived from how often the master data repository is synchronized. Unless the repository is updated in real time it is not recommended to model attributes which have values depending on business logic from source applications.

4.4.8.1 Metadata Model

When speaking about metadata model in connection to product data, it is important to keep the model minimized even though product data sets tend to be of great volume of attributes. In order to add tags to metadata it is crucial to understand the source data. Typical examples of metadata are:

- Source system ID, name and description, function
- Attribute name, data type

Outcomes from this activity serve as a foundation for activities:

(1) Assign Standardization and Cleansing Rules
(2) Assign Matching Rules

4.4.9 Identify Input Interfaces

When importing data from both internal and external sources, the key role play the outcomes from activity Profile and Analyze Data.

Product data is of various source structures in the enterprise, it is therefore important to have a vital metadata management to tag all sources at the input. Product data is also specific by being in various binary files (documents, catalogs, forms, pictures and other) from which it needs to be extracted (depending on if they are machine or human readable).

Based on data analysis and underlying analysis of source system processes, the source attributes and entities are mapped onto the solutions attributes and entities.

There are several types of importing data based on various characteristics:

- Importing product data in terms of synchronization frequency:
  - Batch – loading the data in files transferred from the source system in specified periods
  - Real time – transferring the data in real time or almost real time
- Importing product data in terms of data amount:
  - Incremental – containing only changed or added records
  - Full load – containing all records in the source
- Importing product data in terms of automation (in the software application it is defined how the workflow for data loading is set):
  - Manual import
  - Automated import

Input interfaces vary on the enterprise. Examples of input interfaces can be found below:

- Finance/Accounting
- Material data from suppliers (catalogs, tables, etc.)
- Web content management systems

Outcomes from this activity serve as a foundation for an activity:

(1) Assign Standardization and Cleansing Rules

4.4.10 Assign Standardization and Cleansing Rules

As mentioned in previous chapters, product data is very different from customer data and need a very different and complex approach. This chapter contains a list of product characteristics and their impact on cleaning product data. Nevertheless it does not intend to provide with a complex set of rules to clean product data, as they always depend on the data source, context of enterprise data quality and company culture.

As the author had no direct access to product data, the main resource for best practices for product data cleansing specifics were taken from the former chapters of this thesis, work experience of the author and the following studies:

- Three Critical Steps to Improving Product Data Quality[15] [HAR2010]
- Oracle Product Data Quality [ORA2010]
- Product Data Quality – Different Problem, Different Solution [BOY2009]

This activity is based on outcomes from activity Profile and Analyze Data, without thoroughly understanding the enterprise data, no standardization and cleansing rules can be applied.

There are several views of data quality, depending on the author; each has different metrics and aspects. For example David Loshin recognizes the following data quality dimensions: uniqueness, accuracy, consistency, completeness, timeliness, currency, conformance and referential integrity. [LOS2006]

It is important to involve data quality metrics when making a data quality assessment and monitoring.

Each data quality initiative requires a systematic and ongoing approach. Therefore some data quality checks and monitoring should be set when implementing the solution and continuously run even after the PIM adoption has come to an end to observe the quality of data within the enterprise.

When dealing with data quality issues, the approach typically requires:

[15] Jim Harris has over 20 years of experience in data quality and its related disciplines

- Validations based data type and various rules
- Validations against internal or external reference data tables
- Enrichment from internal or external reference data tables

When using external reference data tables, they need to be acquired from the external institution (activity Acquire External Reference Data). More information can be found in the chapter Product Category. The appropriate cleansing rules are assigned based on the internal structure of the specific table.

The following list contains frequent data quality aspects and challenges of product data in terms of cleansing and standardization rules:

- Categories and hierarchies – different rules for each product category
- Poorly structured data
- Infinite variability – format, content (no fixed syntax, few standards)
- Cleansing product codes, brands, makes, models, catalog numbers, etc.
- Converting numbers of product content
- Packaging volumes
- Creating consistency across multiple suppliers/vendors
- Product variances (different types of the same product) – product series
- Translating values (in case the enterprise has partners or subsidiaries abroad)

Product data is specific because of big volumes of categories, as was mentioned in the third chapter. Standardizing product data is complex as each category has a specific set of attributes, terms, abbreviations, valid values and therefore it must also have unique sets of cleansing rules (Figure 6 shows numbers of categories in enterprises). Therefore it should be considered to which point it is efficient to create standardization and cleansing rules manually for each category and when to use semantic approach. Some enterprises use manual approach regardless of the number of categories. [BOY2009]

When creating standardization and cleansing rules, the above mentioned product data quality aspects need to be taken into consideration. To give a few examples, when identifying a certain product, the packaging plays a very important role, both regarding the number of items in the packaging as well as the weight or content information. Since the brand and product name do not change in case of different packaging (but of an identical product name), it is important to

appropriately locate and cleanse the product content or weight while keeping various measuring standards in mind.

Another aspect of categories and hierarchies is that it brings another level of identification in terms of grouping, opposed to address or party data. This means, that a product may be identified as a part of a certain category or a hierarchy.

Examples of product data can be found in Appendix 10.7.

Outcomes from this activity serve as a foundation for an activity:

(1) Assign Matching Rules

4.4.11 Acquire External Reference Data

Typically external reference data is used either for verification or enrichment of the enterprise data.

There are two reasons for acquiring external reference data related to product data:

- International identification of products
- Need of compliance with regulations

Typical examples for selecting an external reference data table are typically:

- Which enterprise business partners use an international standard (identification)
- Need of regulatory compliance due to importing products from abroad

Depending on the standard, in some cases the data sources for both may be found in one external reference data table as some data contains both types of information.

The follow-up activity is assigning the reference data with appropriate cleansing rules in order to use the external reference data for enriching or validating the enterprise product data.

Outcomes from this activity serve as a foundation for the follow-up activity:

(1) Assign Standardization and Cleansing Rules

4.4.12 Assign Matching Rules

After the enterprise data has been standardized and cleansed, product records must be consolidated using matching rules. This is done upon comparing product records using various attribute sets distinctive for the given entity. The aim of consolidation is to connect duplicated product records from several data sources, which correspond to the same product, and select the best attribute values for their master record.

For this activity the same resources were used as for activity Assign Standardization and Cleansing Rules in combination with findings in the previous chapters and work experience of the author.

The following list contains frequent data quality aspects and challenges of product data in terms of matching rules:

- Product categories
- Product variances (different types of the same product) – product series
- Different sizes of the same product
- Converting numbers of product content
- Packaging volumes

Jim Harris does not recommend, that records are matched only in case of an exact match [HAR2010]. On the other hand, there may be very small variabilities between products, such as attributes from the same series or seasonal special offers of slightly modified products. Another example when this approach should be applied carefully is matching products of various content due to wrong understanding of its volume (250ml is the same as 0,25l, but 0,5l is not 5l). This may sound obvious for humans, but can cause difficulties to the data quality tool.

As a result, a compromise should be established in order to not consolidate different products and at the same time consolidate duplicated products so that there would not be a minimum or preferably no duplicities in the repository.

4.4.13 Identify Output Interfaces

Consumers of the product master data were identified in the activity Identify Consumers of Master Data. Each consumer typically has different requirements for both specific information and synchronization frequency.

This is why each interface is going to require a different attribute set of data depending on which data the consumer uses and if the consumer has any application specific data. The data types of the selected attribute sets may also vary depending on consumer data types, therefore the values may be transferred into the desired data types on the output (e.g. transferring date into varchar data type). The required values are in the end mapped to output attributes (attributes on master level may be split into two output attributes or connected into a single output attribute).

Publishing periods will also need to be set as they may be different for each consumer. Some publishing activities may be automated (either real time or based on a scheduler or event), some may be run only upon a manual interference.

Product master data may or may not be automatically propagated into the source systems. An alternative to propagating the values automatically is a manual rewrite of the source data by a data quality steward or a combination of both options depending on the severity of the data correction.

Output interfaces for both internal and external data consumers are generally called views. Typical examples of master data consumers may be found below:

(1) Internal
- Catalog publishing
- Customer Support
- Risk
- Marketing
- Internal operations

(2) External
- Trading partners
- Audit firms
- Government

Types of export correspond to the same types of importing data.

4.5 Chapter Summary

This chapter contains a general methodology for PIM adoption. It is based on the theoretical findings from previous chapters of this thesis and from personal work experience of the author.

In the first part the three main components of PIM adoption are outlined as well as their functions for better understanding of the concept. All components are further explained and described in separate chapters.

The whole methodology is created regardless of the technological solution and therefore is applicable on any software tool. This is important to keep in mind and apply the methodology in context of the specific requirements of each PIM solution and application software specifics.

5 Conclusion

This work provides an overview of the problematics of the Master Data Management field, both theoretical and practical, with a specific focus on Product Data domain.

The shown complexity of the topic requires both enterprises and external MDM consultancies to interact closely when implementing a Master Data Management solution. Each involved party needs to gain sufficient knowledge of unique enterprise processes and problems as well as requirements given by company, industry or public authorities. Best practices and known software limitations need to be identified and communicated. Each MDM solution must be tailor made for the given enterprise's characteristics and needs.

Contrary to Customer Data, Product Data has a huge diversity. While classifications of products exist on the basis of internal enterprise processes, there is no generic way to categorize products due to the variability of products that could be used or recommended.

To provide an impression for both enterprises considering implementing Product MDM and vendors to embark on Product MDM, the current situation of Product Data in companies was described and problems that exist in terms of data quality and internal structures were elaborated on. Available tools on the market and suppliers specialized on different industry sectors were investigated in order to help interested parties make choices.

Finally, a methodology for implementing Master Data Management of Product Data was provided which can be used as a reference for both enterprises and MDM vendors.

However Product MDM must be seen as a continuous task as each MDM solution needs to be properly adapted to the everchanging enterprise structure, its processes and its decisions on products.

6 References

6.1 Publications

[DRE2008] DREIBELBIS, Allen, et al. *Enterprise Master Data Management: an SOA Approach to Managing Core Information.* IBM Press, 2008. ISBN 978-0-13-236625-0.

[HUG2011] HUGOS, Michael. *Essentials of Supply Chain Management.* John Wiley & Sons, 2011. ISBN 978-0-47-094218-5.

[LOS2009] LOSHIN, David. *Master data management*. Morgan Kaufmann OMG Press, 2009. ISBN 978-0-12-374225-4.

[FEU2008] FEUERLICHT, Jiří. *Enterprise Computing: Concepts, standards and architecture.* Vysoká škola ekonomická v Praze Nakladatelství Oeconomica, 2008. ISBN 978-80-245-1367-6.

[STA2011] STARK, John. *Product Lifecycle Management: 21st Century Paradigm for Product Realisation.* Springer London, 2011. ISBN 978-0-85729-546-0.

6.2 Studies, Articles, White Papers

[BOY2009] BOYD, Martin. *Product Data Quality – Different Problem, Different Solution.*2009. [online] [cit. 2012-12-08] Retrieved from: <http://mitiq.mit.edu/IQIS/Documents/CDOIQS_200977/Papers/02_08_2D-2.pdf>

[GAR2010] Gartner Group Study. *Magic Quadrant for Master Data Management of Product Data.* 2010.

[GAR2011] Gartner Group Study. *Magic Quadrant for Master Data Management of Product Data.* 2011.

[GAR2012] Gartner Group Announcement. *Gartner Says Master Data Management Software Revenue to Grow 21 Percent in 2012.* 2012. [online] [cit. 2012-10-19] Retrieved from: <http://www.gartner.com/it/page.jsp?id=1886314>.

[GLA2000] GLASGOW, Bo, & VAN ARNUM, Patricia. *The time has come: ERP vendors strive for profits*, Chemical Market Reporter. 2000, vol. 258, nr. 5, pp. FR3-FR6. [online] [cit. 2012-10-21] Retrieved from <*http://search.proquest.com/docview/194722091?accountid=17203*>.

[HAR2010] HARRIS, Jim. *Three Critical Steps to Improving Product Data Quality*. 2010. [online] [cit. 2012-11-20] Retrieved from: <*http://www.dataflux.com/ResourceFiles/White-Paper/DataFlux-WP073-Three-Critical-Steps-to-Improving-P.aspx*>.

[IBM2012] IBM Survey: *State of Marketing.* 2012. [online] [cit. 2012-10-16] Retrieved from: <*http://public.dhe.ibm.com/software/marketing-solutions/campaigns/marketer-survey/IBM_State_Of_Marketing_Survey_2012.pdf*>.

[INF2009] Information Difference study. *The State of Data Quality Today.* 2009. [online] [cit. 2012-11-23] Retrieved from: <*http://www.pbsoftware.eu/uk/files/download/analyst-reports/TheStateofDataQualityTodaybyInformationDifference.pdf*>.

[MOT2011] Motorola Survey: Holiday Shopping Survey. 2011. [online] [cit. 2012-10-08] Retrieved from: <*http://www.motorola.com/web/Business/_Documents/static%20files/2011_Holiday_Shopping_Survey_North-America.pdf*>.

[NEM2012] National Electrical Manufacturers Association: The Impact of Regulatory Costs on Small Businesses. 2012. [online] [cit. 2012-12-03] Retrieved from: <*http://www.nema.org/Policy/Documents/NEMA%20Statement%20on%20Regulatory%20Reform.pdf*>.

[ORA2010] Oracle Data Sheet: *Oracle Product Data Quality.* 2010. [online] [cit. 2012-11-20] Retrieved from: <*http://www.oracle.com/us/products/applications/master-data-management/057505.pdf*>.

[POW2009] POWER, Dan. *Product Information Challenge*. Information Management. 2009, vol. 19, nr. 4, pp. 30-31. ISSN 15212912. [online] [cit. 2012-09-18] Retrieved from: <*http://search.proquest.com/docview/214685695?accountid=17203*>.

[POW2010] POWER, Dan. *Why Product Information Management?* Information Management. 2010, vol. 20, no. 3, pp. 43. [online] [cit. 2012-09-18] Retrieved from: <*http://search.proquest.com.ezproxy.vse.cz/docview/356940226?accountid=17203*>.

[ROW2011] ROWE, Nathaniel. *From Design to Delivery: Enhancing the Product Lifecycle with Master Data*. 2011. [online] [cit. 2012-10-10] Retrieved from: <*http://liaison.com/docs/analyst-reports/aberdeen---enhancing-the-product-lifecycle-with-master-data.pdf*>.

[RUS2012] RUSSOM, Philip. *TDWI Best Practices Report: Next Generation Master Data Management.* 2012. [online] [cit. 2012-10-11] Retrieved from: <*http://public.dhe.ibm.com/common/ssi/ecm/en/iml14312usen/IML14312USEN.PDF*>.

[VEN2011] Ventana Research Study. *Value Index for Product Information Management*. 2011. [online] [cit. 2012-09-12] Retrieved from: <*http://www.stibosystems.de/Admin/Public/DWSDownload.aspx?File=%2FFiles%2FFiler%2Fpdf%2FVentana_2011_PIM_Value+Index_Stibo_Systems.pdf*>.

[VEN2012a] Ventana Research Study. *Product Information Management*. 2012. [online] [cit. 2012-09-12] Retrieved from: <*www.ventanaresearch.com/WorkArea/DownloadAsset.aspx?id=3251*>.

[VEN2012a] Ventana Research. *Trends in Product Information Management*. 2012. [online] [cit. 2012-10-06] Retrieved from: <*http://www.ventanaresearch.com/uploadedFiles/Content/Landing_Pages/Ventana_Research_Product_Information_Management_Benchmark_Research_Presentation.pdf*>.

6.3 Internet Sources

[NET_01] International Standard Serial Number – Official Website of the Standard <*http://www.issn.org/*>

[NET_02] Adastra Corporation on Master Data Management <*http://www.adastra.cz/en/990_master-data-management.aspx*>

[NET_03] Harmonized System – Official Website of the Product Classification Standard <*http://www.allhscodes.com/*>

[NET_04] eCl@ss – Official Website of the International Product Classification Standard <*http://eclass.eu/*>

[NET_05] The United Nations Standard Products and Services Code – Official Website <*http://www.unspsc.org/*>

[NET_06] TARIC code – Official Website of the Product Classification Standard <*http://ec.europa.eu/taxation_customs/dds2/taric/taric_consultation.jsp*>

[NET_07] IMDS standard – Official Website of the Product Classification Standard <*http://www.mdsystem.com*>

[NET_08] UN classifications Registry – Official Website <*http://unstats.un.org/unsd/cr/registry/*>

[NET_09] Global Product Classification – Official Website of the Product Standard <*http://www.gs1.org/gdsn/gpc*>

[NET_10] Kompass Classification – Official Website of the Product Standard <*http://www.kompass.com/*>

[NET_11] Software and Systems Engineering Vocabulary <*http://pascal.computer.org/sev_display/index.action*>

[NET_12] Industry Classification Benchmark – Official Website <*http://www.icbenchmark.com/*>

[NET_13] International Standard Book Number – Official Website of the Standard <*http://www.isbn.org*>

7 List of Figures

8 List of Tables

9 Glossary

Expression	Abbreviation	Definition
Bill of Material	BOM	Documented formal hierarchical tabulation of the physical assemblies, subassemblies, and components needed to fabricate a product [NET_11]
Customer Data Integration	CDI	MDM Systems that focus exclusively on managing information about customers. [DRE2008]
Data Quality	DQ	Deals with data validation and cleansing services (to ensure relevance, validity, accuracy, and consistency of the master data), reconciliation services (aimed at helping cleanse the master data of duplicates as part of consistency), and cross-reference services (to help with matching master data across multiple systems). [DRE2008]
Enterprise Resource Planning	ERP	Enterprise computing applications typically consisting of a large number of modules that support various business functions within the enterprise, e.g. financials, order management, etc. [FEU2008]
Master data	-	Master data is the core information for an enterprise, such as information about customers or products, accounts or locations, and the relationships between them. In many companies, this master data is unmanaged and can be found in many, overlapping systems and is often of unknown quality. [DRE2008]
Master Data Management	MDM	Through a combination of architecture, technology, and business processes, MDM provides an approach to incrementally reducing the amount of redundantly managed information and providing information consumers throughout an enterprise with authoritative master data. [DRE2008]
MDM Solution	-	An MDM Solution addresses the MDM-related problems and challenges that exist in a specific customer situation; it is comprised of those MDM components that need to be complemented mainly with an MDM strategy, a set of architectures, products and technologies, and best practices. [DRE2008]
Master Data Management of Customer Data	-	See Customer Data Integration.

Master Data Management of Product Data	-	See Product Information Management.
Product data	-	Information about products such as dimensions, weight, etc. [AUTHOR]
Product Information Management	PIM	MDM Systems that focus exclusively on managing the descriptions of products. [DRE2008]
Product Lifecycle Management	PLM / PLCM	Product Lifecycle Management is the business activity of managing, in the most effective way, a company's products all the way across their lifecycles; from the very first idea for a product all the way through until it's retired and disposed of. [STA2011]
Structured data	-	Data which is organized in a structure, hence it is easier to be identified and understood by computers, e.g. database or an xml file. [AUTHOR]
Unstructured data	-	Data with no identified structure, although there may be some sort of loose structure typical for the specific type, e.g. pictures, documents, videos. [AUTHOR]

10 Appendices

10.1 Vendors Specialized on PIM

Name	URL	Product Name[16]
ADAM	http://www.adamsoftware.net/	ADAM PIMS
Advanced Concepts	http://www.advanced-concepts.de/	ac.PIM ac.ePIM
Agility	http://www.agilitymultichannel.com/	
CaTS	http://www.systemintegration.net/	
DataFlux[17]	http://www.dataflux.com/	qMDM
Data Ladder	http://www.dataladder.com/	
Datactics	http://www.datactics.com/	
EnterWorks	http://www.enterworks.com/	
Epitomy	http://www.epitomy.com/	
Geni-Sys	http://www.geni-sys.net/	
GXS	http://www.gxs.com/	PIM for Retail Edition PIM for Supplier Edition
Heiler Software	http://www.heiler.com/	Heiler Product Manager
Hybris	http://www.hybris.com/	Hybris PCM
IBM	http://www.ibm.com/	MDM Server for PIM
Informatica	http://www.informatica.com/	Informatica MDM
jCatalog	http://www.jcatalog.com	
Oracle	http://www.oracle.com/	Oracle PIM

[16] In case a value is missing in Product Name column, it is either because the company has no specific name for its product, because it provides PIM functionality within a different company product or because several SW applications are needed to cover MDM of Product Data

[17] Data Flux has been acquired by SAS, for this reason SAS is not listed separately but is in this case included under this entry.

Orchestra Networks	http://www.orchestranetworks.com/	EBX
Perfion	http://www.perfion.com/	Perfion PIM Solution
QAD	http://www.qad.com/	
Riversand	http://www.riversand.com/	Riversand MDMCenter
SAP	http://www.sap.com/	SAP NetWeaver MDM
Sepia	http://www.sepia.de/	Sepia Alterra PIM
Stibo Systems	http://www.stibosystems.com/	STEP MDM
Tibco Software	http://www.tibco.com/	Tibco Collaborative Information Manager
Trading Bell	http://www.tradingbell.com/	CatalogStudio PIM
Viamedici	http://www.viamedici.com/	Viamedici EPIM
WebOn	http://www.webon.no/	WebOn PIM Solution
Zynapse	http://www.zynapse.com/	Zynapse MMDM

10.2 Example of a Product Master Data Table in the Pharmaceutical Industry

Attribute Name	Description	Attribute Type
Product Master ID	Unique ID	integer
Brand Name	Drug brand	varchar(100)
Description Field	Drug description	varchar(1000)
Drug Identification Number	Internal ID of drug numbering	varchar(100)
Drug Supplier	ID of drug supplier	varchar(100)
Manufacturer	ID of drug manufacturer	varchar(100)
Country of Origin	Country of production	varchar(100)
Year	Year when the drug went into production	varchar(4)
Drug Status	Internal drug status	varchar (10)
Status Date	Date when Drug Status was set	date
Class	Internal classification of the drug	varchar(100)
Active Ingredient(s)	Active ingredient(s) in the drug (international standard)	varchar(1000)
Active Ingredient Group (AIG) Number	Group number of active ingredient contained in the drug	varchar(100)
Active Pharmaceutical Ingredient (API)	Substance in the drug	varchar(100)
Strength	Internal company rating of drug strength (from 1 to 100)	integer
Pharmaceutical Form(s)	Unstructured data containing internal forms	blob
Product Monograph (PM)	Monograph provided by a pharmaceutical manufacturer	varchar(100)
Therapeutic Classification (ATC)	International drug classification	varchar(10)

Bar Code	Bar code representation of EAN-13	varchar(20)
Drug Volume	Content (number of pills/liquid volume)	integer
Drug Weight	Weight in grams	integer
Discontinued	Boolean value if product is currently being sold	varchar(1)
Create Date	System attribute expressing the date when master record was created	date
Created by User	System attribute expressing which user created the master record	varchar(10)
Update Date	System attribute expressing the date when the master record was updated	date
Updated by User	System attribute expressing which user updated the master record	varchar(10)

10.3 Example of Internal Reference Tables in the Pharmaceutical Industry

10.3.1 Drug Status on Master Level

Status Code	Description
0	Drug was discontinued because it was not manufactured
1	Drug was discontinued but is still manufactured
2	Drug is currently being sold
3	Drug has been put into inventory but is not sold yet

10.3.2 Drug Status on Reference Level

In the table below is depicted an internal reference data table connected from two source systems (System A, System B) with the value represented in master table (see above). As is showed, the same meaning on master level may be represented by two various status codes in the source systems.

Source system	Status Code	Master Status Code	Description
System A	A	2	Drug is currently being sold
System A	N	0	Drug is not produced any longer
System B	0	0	Drug was discontinued because it was not manufactured
System B	1	1	Drug was discontinued but is still manufactured
System B	2	2	Drug is currently being sold
System B	4	3	Drug has been put into inventory but is not sold yet

10.4 Example of an Internal Reference Data Table in the Financial Sector

Below is an example of an internal reference table in the financial sector:

10.4.1 Product Type

Product Code	Description
10	Deposit product
11	Insurance product
21	Loan product
31	Interest rate product
32	Equity and security product (stocks, bonds)

10.5 Example of External Reference Data Table

10.5.1 Central Product Classification (CPC)

The table below lists the top level of Central Product Classification:

First Level Code	Description
0	Agriculture, forestry and fishery products
1	Ores and minerals; electricity, gas and water
2	Food products, beverages and tobacco; textiles, apparel and leather products
3	Other transportable goods, except metal products, machinery and equipment
4	Metal products, machinery and equipment
5	Constructions and construction services
6	Distributive trade services; accommodation, food and beverage serving services; transport services; and electricity, gas and water distribution services
7	Financial and related services; real estate services; and rental and leasing services
8	Business and production services
9	Community, social and personal services

10.6 Example of Classification Standards for Drinking Glasses

Example of classes for glassware (drinking glasses) for the following standards and the comparison of different structures:

10.6.1 Central Product Classification

Code	**Description**
First Level Code	**Description**
3	Other transportable goods, except metal products, machinery and equipment
Second Level Code	**Description**
37	Glass and glass products and other non-metallic products n.e.c.
Third Level Code	**Description**
371	Glass and glass products
Fourth Level Code	**Description**
3719	Other glass articles
Fifth Level Code	**Description**
37193	Glassware of a kind used for table, kitchen, toilet, office, indoor decoration or similar purposes (except bottles, jars and the like, of glass, and ornaments of lamp-worked glass)

10.6.2 Harmonized System

Example of Harmonized System categorization shows data quality issues in the description column (taken from the official website [NET_03]).

Code	**Description**
First Level Code	**Description**
7013	Glassware of a kind used for table, kitchen, toilet, office, indoor decoration or similar purposes (other than that of heading 70.10 or 70.18).
Second Level Code	**Description**

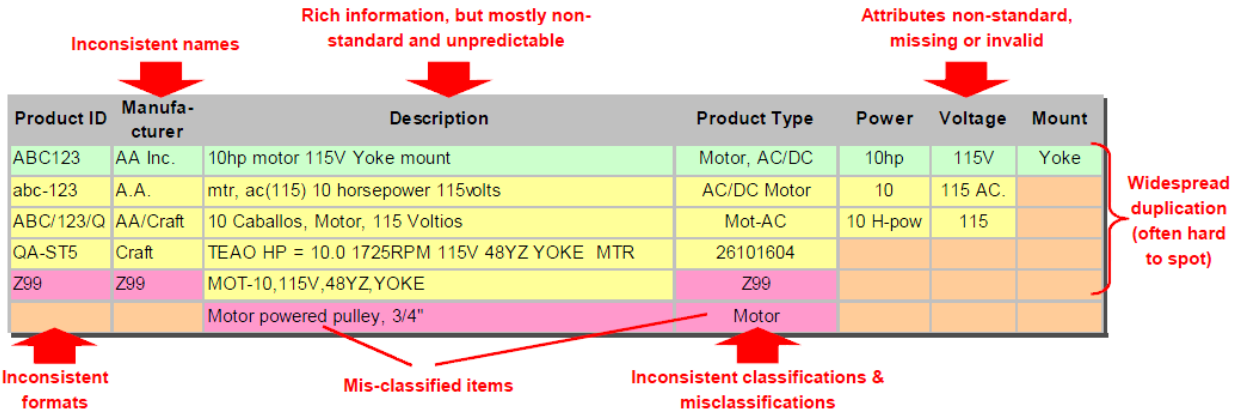
701310	Glass-Ceramic Objects, Used For Table, Kitchen..., Nes
701322	Glassware...Stemware drinking glasses other than of glass-ceramics of lead crys
701328	Glassware...Stemware drinking glasses other than of glass-ceramics Excl. of lea
701333	Glassware..Other drinking glasses, other than of glass ceramics Of lead crysta
701337	Glassware..Other drinking glasses, other than of glass ceramics Excl. Of lead
701341	Glassware of a kind used for table (other than drinking glasses) Of lead crysta
701342	Glassware..(othr than drinking glasses)Of glass having a linear co-effi..=<5x10
701349	Glassware ...(othr than dri.glasses) or kit.purposes other than of glass-cerami

10.6.3 eCl@ss

Code	**Description**
First Level Code	**Description**
29	Home economics, Home technology
Second Level Code	**Description**
29-09	Dishes, cutlery, glassware
Third Level Code	**Description**
29-09-01	Glassware
Fourth Level Code	**Description**
29-09-01-90	Glassware (unspecified)

10.7 Example of Common Product Data Problems

The table below shows an example of problems typically found in product data from [ORA2010]:



The following example of common product data problems is taken from [HAR2010]:

Category	Brand	Product Description
Candy	E<3MC^2	Candy Bar Milk Chocolate Square Net Weight 3.5 oz.
Candy	Everybody Loves Milk Chocolate Squared	NET WGT 99G Candy Bar
Candy	Milk Chocolate Square that Everybody Loves	in a 3.5 oz Candy Bar
Candy	e-heart milk chocolate squares	35oz box of chocolate candy bars
Candy	(ELMC2)	Candy Square Bar 99g of Milk Chocolatey Goodness
Candy	E-Heart Emoticon-Milk-Chocolate-Squared	3-and-1/2 ounce BAR
Candy	E3MC^2	Square Chocolate Candy Bar 99g (3.5z) Milk Chocolate
Candy	Milk Chocolate Squares Everyone Loves	10 1.5 OZ Squares (15 OZ BAG)
Candy	Sugary X-treme	Non-Sugar-Free Chewing Gum Net Weight 35 grams
Candy	Sugar Extreme	Six Pack of Sugar Chewing Gum (Net Weight 7.41 oz.)