Product Information Management: Definition, Purpose, and Offering

SapientNitro℠

Product Information Management: Definition, Purpose, and Offering

Product Information Management (PIM) solutions create a single view of a product for an enterprise. The single view of a product can be used for both operational and analytical purposes. PIM solutions serve information on demand—independently of any other repository of product data—for any designated destination, be it Web site; a mobile application; a kiosk at a store; or a publication (paper, CD-ROM, or online).

Why Use a PIM Solution?

In 2010 Stibo reported the following survey result: 44% of companies still use spreadsheets to manage and share product data. This manual process is typically cumbersome and leads to poor, inconsistent, and incomplete item data: a product is not described the same way across the enterprise as users access product information through various information systems. Although product vocabularies may exist, they are often not enforced. At the end of the day, companies have product descriptions that are ambiguous, incomplete, useless, inconsistent, or inaccurate.

If fields are not documented, it leads to poor end-user experience on a retail Web site whenever the customer uses guided navigation. For a case in point, consider what happens when a user visits an electronics category in which 50% of the products do not have the brand documented. If she uses brand as a filter, 50% of the retailer's products will be missing from her search results. This represents a lost opportunity to sell products.

When the quality of the item information is poor, it leads to page abandonment and cart abandonment: it is always cheap and easy for a customer to go to another site—where she will have more complete information during the process of purchasing. This, in turn, will lead to a lower conversion rate.

Data inaccuracy has far-reaching effects downstream throughout the enterprise. For example, U.S. Xpress, the third largest privately-owned trucking company in the United States, reported in CIO.com how a concern for fuel cost associated with truck idling times prompted them to engage in an information management strategy: after

reassessing a very disparate IT infrastructure composed of 130 different applications—with no data integration, stewardship for the data, or master database—U.S. Xpress was able to save about $6 million a year across its fleet of 8,000 tractors and 22,000 trailers.

The spreadsheet-based process also leads to increased labor costs. Retailers often complain about the labor costs associated with setting up product information. These costs arise from rework and from the complexity of figuring out where the information is coming from, which sources to trust, and how to share the information with other people. The complexity and stress of figuring these things out also lead to a high turnover rate for the staff on the retailer's item setting team, as well as strained relations between the retailer and its suppliers.

Finally, the spreadsheet-based process increases a product's time-to-market (i.e., the time it takes to collect all of the product information and publish it on the site). An increase in time-to-market can be seen as a labor cost *or* as an opportunity cost in not making the product available earlier on the site for sale.

Master Data Management (MDM) and PIM Solutions

Master Data is high-value core information that exists across an enterprise and is used to support critical business processes. It is typically related to customers, suppliers, vendors, partners, and products. (**Note**: Master Data is a subset of your enterprise data; it is not *all* of your enterprise data.) Master Data is at the heart of every business transaction, application, and decision—whether it involves selling a product or interacting with a customer.

PIM is MDM applied to the Product Information domain. MDM is a *business capability* for ensuring that an organization does not use multiple (potentially inconsistent) versions of the same Master Data in different parts of its operations. (Note the italics in the previous sentence: MDM is a *business capability*, not a technology solution.) The need for MDM comes from the plethora of systems—nine on average, according to a 2010 survey by Information Difference—that generate product Master Data. MDM is enabled through alignment of multiple information technologies, business process improvements, and organizational commitments.

Short of having the above three components—alignment of multiple information technologies, business process improvements, and organizational commitments—you do not have MDM. In those instances, an MDM pure-technology solution, whether it is a PIM solution or something else, will not solve the problem. You have to look at your business process and your information supply chain, and you have to look at how you are going to govern that process and that data. Who is going to decide which attributes you are going to collect for a product category? Who is going to say that the Brand attribute, in Electronics, is mandatory in order to sell the product? Who is going to understand the impact of mandating that information on your supply chain and on your vendors?

Commonly Used MDM Terms

• **Global Data Synchronization Network (GDSN)** is an international communication protocol led by GS1, that enables companies around the globe to exchange standardized and synchronized supply chain data with their trading partners (e.g., between manufacturers, distributors, and retailers). The primary focus of GSDN is on the supply chain, especially dimensions—not just the assembled dimensions of the product—but the inner pack, the pack, the box, the number of boxes per pallet, the UPC for the box, for the pallet, for the pack and inner pack, and so on and so forth. There is less emphasis on the faceted contents such as the rich attribution associated with a category.

• **Global Product Classification (GPC)** is a GS1 standard for the purpose of categorizing products. All of us have our own concepts of categorization. With GPC there is a shared view of categorization.

• **Data Governance (DG)** is the formal orchestration of people, processes, and technology across an organization to enable the leveraging of data as an enterprise asset. DG is typically put in place when an organization finds out that data is an asset in the same way that factories, workers, and knowledge are assets.

• **Customer Data Integration (CDI)** is the equivalent of PIM but is focused on customer information. It consists of processes and technologies for recognizing a customer and their relationships at any touch-point, while aggregating, managing, and harmonizing accurate, consistent, complete, and up-to-date knowledge about that customer to deliver it just in time in an actionable form to touch-points.

• **Multi-Entity MDM** is an MDM solution to concurrently manage multiple, diverse Master Data domains (customers, accounts, products) across intra- and extra-enterprise business processes. The idea behind Multi-Entity MDM is to build MDM that cuts across categories, whether they are products, vendors, locations, or customers.

Business Drivers for PIM Implementation

The primary business driver for PIM implementation is the retailer's need to shorten the New Product Introduction (NPI) cycle. A retailer's motive for shortening the NPI cycle is simple: by doing so she gets more time to sell the product. A PIM solution can help to reduce the NPI cycle by enabling the retailer to do the following:

• Reconsider and redefine its unique, proprietary workflows. This makes it possible to remove human bottlenecks and streamline operations. A Workflow Engine should be part of a PIM solution. It allows the retailer to orchestrate the flow of information associated with products, whether it is collection, validation, or approval. Because of the need to describe the retailer's workflow inside the PIM, the implementation of a PIM provides a great opportunity for the Business Analysts to work with the Client—not only to assess what exists and what the best practices are—but also to investigate where and when they can streamline the flow. An example of a workflow appears in Figure 2 on page 8.

• Implement a metadata repository and associated business rules. This increases accuracy and reduces rework. For example, a PIM will allow you to not only enter the information associated with a product—such as the UPC "051111461860" (which is the data)—but also describe what that information should be, in this case exactly 12 digits, using 0 padding (the metadata). This drives the validation that will occur whenever that data is captured (whether manually or through feeds). An example of a business rule might be that the "Brand" field is mandatory for all products in the "Electronics" category, but not in the "Apparel" category.

• Build an information exchange and implement system-to-system integration. The exchange of product information via a PIM solution is faster and less error-prone than the exchange of information via telephone, instant messaging, or e-mail. Consider what happens when a retailer sends product information—in a spreadsheet attached to an e-mail— to one of its vendors: the retailer hopes that the recipient opens the e-mail, opens the attached spreadsheet, adds data to the spreadsheet, and then sends it back. When the retailer gets the data back, she may realize that the 12-digit UPC—which was originally sent as text—came back as a number which is now in exponential format. The UPC is now useless and the retailer needs to go back to the vendor.

• Provide a supplier portal into the actual item data workload. This simplifies work, increases suppliers' visibility of products that will be carried by the retailer, and augments ownership in the process.

The Web portal gives visibility to the SKUs that the retailer plans to carry and sell. If the SKU count is large, the supplier has the opportunity to increase staffing to provide the information in a timely fashion. If the information is incomplete or does not comply with the validation stated by the metadata or the business rules, the supplier gets immediate feedback through the use of the Web portal. This speeds up the information collection process and increases quality compared to a traditional batch approach (whether emails or ftp) in which data is submitted by the supplier to the retailer, who then attempts to insert that data in the PIM, where validation might fail. If validation fails, errors are then sent back to the suppliers, increasing the number of round trips, delaying time-to-market, and fostering employees' frustration over the lack of visibility.

• Eventually implement a vendor's scorecard for data quality and turn-around. This increases accountability (no product data, no product sale).

Because PIM solutions improve the quality of product information, they improve the operational efficiency of a retailer's supply chain. There are fewer problems in the Upstream Supply Chain (e.g., order fulfillment issues and incorrect Advance Shipment Notices). Also, there are fewer problems downstream (e.g., the customer receiving the wrong product because the UPC was incorrectly captured).

PIM solutions meet the need for the delivery of timely eCommerce initiatives across channels using a single, trusted source of product information.

PIM solutions also allow retailers to manage the Update and Approval process for product information.

PIM Capabilities

A PIM solution allows retailers to do the following:

• Locate and use appropriate product data from heterogeneous sources. For example, while an ERP system might provide the product dimensions, weight, and UPC for the product, it will have none of the rich content associated with that product. The rich content might be provided by a 3rd party content aggregator who will make the information available through a feed that is processed by an ETL (Extract-Transform-Load) process.

• Access structured product data, which consists of such things as the dimensions, weight, name, UPC, description, and features set. Unstructured data, on the other hand, consists of any supplemental material related to the product that is not easily modeled in the PIM repository. It can include such things as video clips demonstrating the product, product warranty (PDF), or rebate form (PDF).

• Cleanse data and related content
    o Very often the product data from a vendor is only a brief paragraph describing the product features. While all of the features may be described, the content has not been curated, processed, and normalized.

o The process of data cleansing ensures that content is consistent across product categories—i.e., uses standardized vocabularies, normalized value lists, the correct style guide, and the correct copy guidelines—and reflects the demography of their shoppers' interest, as well as the value associated with the retail brand.

o A PIM allows us to cleanse the data and associate one item of data with another. For example, for a UPC, data cleansing might ensure that the system complements the UPC with leading 0's so that it has 12-digits.

• Identify and create missing product information. For example, a mandatory "Brand" field might have been present in the initial "Long Description," but the field itself might not have been documented. A PIM validation workflow would allow the retailer to flag those products (in this specific category, with this specific missing information) and draw attention to missing content prior to the product appearing on the Web site.

• Connect and transmit the data

• Unify and relate a single product instance to multiple types of content (i.e., provide unification and a single source of product information for multiple types of content.) By collecting, validating, curating, and approving the product-related content, the PIM provides one synthetic representation, which is available for different purposes.

• Distribute product information from a single source

• Enable cross-media publishing of product catalogs (i.e., product information)

• Create personalized catalog views of the product information. Such a view contains only the product information that the specific user cares about. For example, when a warehouse worker receives a computer, he does not care about the number of gigabytes on the computer's hard drive. Nor does he care about the computer's processor speed. What he cares about is the size of the box, the weight of the box, and whether the box can be placed on a conveyor belt. The PIM solution allows the creation of a customized view for the warehouse worker—that shows him only the product properties that he cares

about—out of the 100 to 200 properties that are associated with each product.

• Enable multilingual catalog creation and deployment. You need ways of presenting that information and rolling it out in workflows, and you will translate that information later for different markets.

• Support integration with Digital Asset Management systems (DAM), which allows retailers to integrate the product information with other digital assets (e.g., digital pictures).
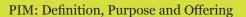
Core Functionality of a PIM Solution

A PIM allows you to create lots of metadata, including descriptions of product categories, descriptions of the information you want to collect, the rules about that information, and the exceptions to those rules.

In addition to the ability to describe the metadata, the core functionality includes the ability to describe the validation rules that are to be enforced during your workflow and when those validation rules should be applicable in the workflow. For example, at the onset of the New Product Introduction workflow, you may not need the brand for an electronic product. However, you will definitely need the UPC, a short description, the dimensions, and assurance that the information will be sent as quickly as possible to the supply chain. (The latter is necessary because it takes time for products to flow through the warehouse.) Meanwhile the copywriter is going to generate the rich copy for the product, and she can research all of the supplemental facts about the product. As a later step, the workflow will enforce the by-now documented brand, which is mandatory for Electronics.

Other core functionalities of a PIM solution include the following:

• Data models to manage product taxonomy, attributes, relationships, and hierarchies

• Internationalization and localization

• Data integration support (both transactional and batch)

• B2B synchronization with trading partners

• Performance and scalability, which enable the PIM solution to support multiple users and large SKU counts

• Data parsing, data cleansing, data standardization, and de-duplication of data

• Data version control with change history

• Workflow and business process modeling—
At the heart of your PIM solution is a Workflow Engine that helps you collect the information at the right time by presenting the relevant data to the relevant employee in the organization or to the vendor (if you externalize your PIM through a portal).

• A PIM solution must include the ability to efficiently search for product or navigate through the product taxonomy, especially considering that it is possible for a large retailer to have more than 4 million SKUs.

• Data governance and stewardship process management

• Linkage of structured product data with unstructured content

• Reporting and auditing

PIM Architecture

There are four PIM solution architectures: External Reference, Registry, Reconciliation Engine, and Transaction Hub.

External Reference (or Consolidation) is the low-end PIM solution architecture. In this architecture, the MDM Data Hub is a reference database that points to all data but does not actually *contain* any data. Unlike the other three architectures (described below), the External Reference architecture does not define, create, or manage a centralized platform where Master Data is integrated to create a "single version of the truth." **Note:** The External Reference architecture is no longer supported by most MDM vendors.

The **Registry** architecture consists of a registry of unique master entity identifiers, which are created using identity attributes. An entity resolution service identifies the master entity records that represent the same entity and therefore should be linked. The data-source links that

were used to maintain the attributes are maintained by the Data Hub. One drawback to this architecture is that the correct attribute values cannot be created by the Data Hub in the event that the data is not available in the source systems.

The **Reconciliation Engine** (or Coexistence) architecture is a step up from the Registry architecture.
It harmonizes product Master Data across databases and acts as a central reference point. The Reconciliation Engine architecture is well-suited for large, distributed organizations with multiple data ownerships. This architecture provides synchronization between itself and legacy systems; retailers will often implement it as an intermediate architecture (i.e., after they have outgrown the Registry architecture but are not yet ready for the Transaction Hub architecture described below).

The **Transaction Hub** architecture stores the up-to-date product Master Data with its associated enriched attribute data. It also supports new and legacy transactional and analytical applications, and includes a business service and data integration layer. This architecture is well-suited to companies that need to collect information, cleanse it, build it on the fly, and serve it to other destinations.
It is also well-suited to companies that have a high rate of change in their upstream data sources.
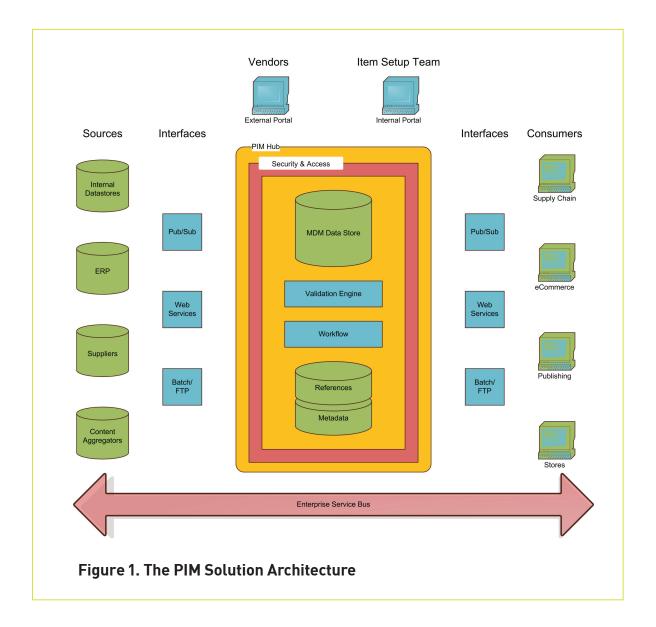
Figure 1. The PIM Solution Architecture

Figure 1 illustrates the general PIM solution architecture. The PIM hub contains the MDM Data Storage, the Validation Engine, the Workflow Engine, References, and the Metadata. This information is made available through the Security and Access Layer. The latter ensures that you present content only to persons who are entitled to have it, even as you allow authorized persons to modify that content. For example, you may want someone in the warehouse to modify the dimensions of a product because she has measured the box. However, you don't want someone in the rich copy team to modify it because she has just *guessed* that it was 2x2x2.

The Security & Access Layer enables authorized users to have customized views (i.e., views that contain a *subset* of the information). The information is made available both upstream and downstream through an Enterprise Service Bus—using mechanisms such as PubSub, Web Services, or Batch FTP—that will allow you to collect the information or publish it to your consumers whether they are supply chain, eCommerce, publishing, or stores.

It is important to allow portals to the product information stores, either for your item setup team (inside the enterprise) or for your vendors (outside of the enterprise). Providing portals to the vendors gives them better visibility in their workload—and better accountability for the quality of their products—because they have direct access to what they are supposed to do and why things are not accepted when they are missing information.

PIM and eCommerce Workflow

Typically, as part of a PIM implementation, you would re-visit the Business Process Workflow
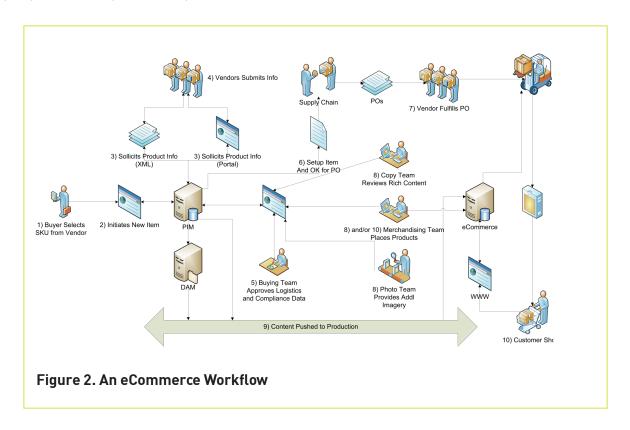

Figure 2. An eCommerce Workflow

Figure 2 contains an example of a workflow for setting up a product in the enterprise. The eCommerce activity is mostly on the right-hand side of the diagram. It depends on upstream activities to collect the needed information (i.e., product name, dimensions, weight, tax and licenses, regulations and compliance, rich content, unstructured content, etc.).

Implementing a PIM Solution

Before you implement an MDM solution—whether it is a PIM solution or some other kind of MDM solution—it is important to have an understanding of what MDM is and is not. The following books provide a good overview of MDM:

• *Master Data Management and Data Governance*, by Alex Berson and Larry Dubov (McGraw-Hill Osborne Media; 2nd Edition; ISBN-13: 978-0071744584)

• *Master Data Management*, by David Loshin (Morgan Kaufmann OMG Press; ISBN-13: 978-0123742254)

The next step is to determine when, where, how, and why MDM should be adopted. This includes evaluating the scope of the MDM project. It is recommended that the first implementation of MDM not encompass the entire organization. For example, in one retail PIM

implementation, it was decided to totally disregard media (i.e., books, music, movies) whether they were in physical or electronic form. This is because there were good processes, automated feeds, automated categorization, and no quality issues with media. Instead the focus was placed on hard lines and soft lines, such as electronics, apparel, video games, toys, and those kinds of products where significant benefit was anticipated.

You need to define the Master Data domains, use cases, implementation styles, organizational structures, technologies, and best practices that you will use in implementing the solution.

Finally, you must analyze the impact that the solution will have on business strategy and business direction and assess which vendors, markets, and industries will interact with the solution.

It is a good idea to adhere to the following roadmap:

• Define and validate the business needs

• Ensure that data governance and business sponsorship are present. The business owner should be sitting side-by-side with the CIO. If both of them are not present, you should be concerned.

• Scope your unique MDM ecosystem

• Analyze your data: if you are going to build one vision of truth, you don't want to migrate all of your legacy information from the system of record without first profiling it, cleansing it, and making sure that it is of reasonable quality.

• Make technology decisions

• Kick off the phased implementation

MDM should not be implemented exclusively as an IT project. Rather, it should be implemented as part of the enterprise vision and direction; in order to have a successful MDM implementation, the business must buy into the project from the outset. It is the business that must take ownership of (and responsibility for) master data governance and stewardship. Some mistakes to avoid in MDM implementation include the following:

• Making the project too large in scope

• Making the MDM effort part of a larger project. This sometimes happens when the MDM and eCommerce replatforming are started concurrently, and the eCommerce project becomes dependent on the success of the MDM project.

• Overanalyzing the requirements (i.e., "boiling the ocean")

• Not establishing governance

• Not modeling the data topic

• Making the MDM implementation purely analytical rather than operational

• Underestimating the political dimension in MDM delivery

PIM Vendors

The dominant PIM solutions are IBM InfoSphere, Oracle Product Hub, and SAP NetWeaver. All three vendors, IBM, Oracle, and SAP, have have been involved with MDM for the past 10 years. They have reached their positions of dominance through multiple acquisitions.

All three vendors offer a full MDM ecosystem, including data integration, data quality, databases, messaging, and sometimes hardware.

One question for each of these vendors is which MDM will they provide at the end of the day? If the vendor is Oracle, do they provide the Siebel solution? If the vendor is IBM, do they provide the Trigo Technologies solution?

While SAP is mainly focused on selling to existing SAP customers, IBM and Oracle sell to both their established customers and new customers.

All of the following vendors have just one MDM solution: Tibco, Riversand, Stibo, Hybris, and Heiler. Except for Tibco, all of these vendors built their solutions in-house. They all feature strong integration with third parties. Riversand, Stibo, and Heiler are MDM pure plays. These vendors have established relationships and built adapters for the most common eCommerce solutions, such as Oracle ATG, IBM WebSphere Commerce, hybris Commerce, and Demandware Commerce.

Talend is an Open Source vendor whose ETL products have been used in various projects at SapientNitro. Although their product is Open Source, they do charge for support. Their MDM solution is the result of an acquisition made last year and has yet to be vetted and proven.

Conclusion

PIM is Master Data Management applied to the product space. PIM is enabled through business process improvements, organizational improvements, and the alignment of multiple information technologies. To successfully implement PIM requires business ownership of the PIM process.

Subject Matter Expert: Christophe Marcant

Christophe Marcant is a Senior Specialist in Sapient's San Francisco office. His focus is on Product Information Management (PIM) systems and product content acquisition. Christophe has over twenty years experience in software development, and he provides technical leadership in all phases of project life-cycle development, from initial feasibility analysis and conceptual design through implementation.

Prior to Sapient, Christophe worked at Walmart Global eCommerce, where he led a multi-disciplinary team that delivered software solutions for Catalog and Content Management. In this capacity Christophe supervised the roll-out of PIM implementations that are still in use at Walmart.com. Christophe has a broad understanding of product data and how to leverage it in an optimal manner.

Christophe holds an M.S. in Computer Science from the Ecole Nationale Supérieure des Mines de Paris and an M.S. in Artificial Intelligence from University of Paris.

© Sapient Corporation 2011