Commentary

How to Manage Product Data in a Multienterprise World

Managing product content and data internally is challenging, and keeping data consistent across suppliers and customers is a Herculean task. Product content and data management is aimed at handling this difficult job.

Product content and data management (PCDM) involves a set of related disciplines, technologies and solutions used to create and maintain consistently interpreted product data and facilitate commercial exchange. PCDM will become increasingly important during the next five to 15 years as enterprises seek to eliminate waste and create competitive advantage (see "Product Content and Data Management Promises Savings"). Although the business drivers for PCDM are clear, the technical underpinnings are not as well-defined. We consider product registries and business-to-business (B2B) data synchronization to be the likely technical responses to the PCDM challenge.

Responding to the PCDM Challenge: The Registry

Because PCDM is a cross-application, cross-departmental, cross-functional and cross-enterprise activity, whole new classes of applications will begin to gain traction during the next five to 10 years. Furthermore, established applications will require extensive overhaul (including a redesign based on new architectures) to contribute to PCDM, and these changes will be leveraged by new supply chain management (SCM) solutions (see "New P2P Solutions Will Redefine the B2B Supply Chain").

For example, typical enterprise applications "assume" that product data persists behind the firewall and that this data "pool" is based on item hierarchies or item numbers. Based on the heterogeneity of product data created by the exchange of information within and between trading communities, this is an insufficient approach. For those industries that seek to create shared external data pools (based on a distributed-based registry model), product data will be anything but static, and it will not be subject to any single enterprise's classification scheme.

In the case of a peer-to-peer (P2P)-based registry model, buyer systems would be required to acquire updates and new product data from external systems. Furthermore, seller systems will have to ensure that external systems are fed and maintained by several internal systems, such as those related to product definition (engineering), sales and marketing, and even logistics systems for item availability. As a result, PCDM systems must support reconciliation with external systems, driving the creation of wholly new applications or services. To maintain consistency in data across enterprises, the following PCDM
services, facilities and capabilities are likely to evolve across industries during the next decade (see Figure 1).

Figure 1
The Keys to Consistent Multienterprise Data

<table>
<thead>
<tr>
<th>Business-to-Business Value-Add Applications and Services</th>
<th>Differentiating Business Processes, Web Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Messaging, Update and Synchronization</td>
<td>Heavy Lifting — TDNs WSNs, EAI, XML</td>
</tr>
<tr>
<td>Pooling, Cleansing and Normalization</td>
<td>Value-Added Service for Gaining Compliance</td>
</tr>
<tr>
<td>Compliance and Endorsement</td>
<td>Independently Managed Standards Endorsed</td>
</tr>
<tr>
<td>Physical Register</td>
<td>Local, Even Independent Web Services for DNS</td>
</tr>
<tr>
<td>Logical Domain Name System (DNS) Register</td>
<td>Common or Shared Foundation for DNS</td>
</tr>
</tbody>
</table>

Source: Gartner Research

B2B Value-Added Applications and Services: These capabilities will be required to sustain the underlying technical infrastructure that is central to registry-based PCDM. In the same way that transaction data has been exchanged for some time via value-added network (VAN) services — such as in electronic data interchange (EDI) VANs — the exchange of product data will be facilitated by service organizations. B2B value-added services will provide the trading community with management and support for variants of industry standards within trading communities. These services are required to support the bidirectional exchange of product data with multiple internal applications, as well as external trading partners.

In addition, traditional enterprise applications — such as enterprise resource planning (ERP) II, supply chain execution (SCE), customer relationship management (CRM), supplier relationship management (SRM), partner life cycle management (PLM) and new applications focused on multienterprise business processes — will evolve to accept product data and change notifications from external enterprises and not just manage the internal processes they were used for previously.

Messaging, Update and Synchronization: A scalable PCDM messaging infrastructure will require bidirectional notification of inaccurate or expired data. Mission-critical buy or sell processes will require real-time processing. As a result, such processes will become a critical requirement for the real-time enterprise (RTE). Notification will be generated from the primary data pools and sent to registered consumers of such data, directly (in the case of P2P networks) or via centralized data registries. These services are fulfilled by enterprise application integration (EAI) vendors, although, as industries consolidate around multiple schema or adopt singular-registry models, these services are likely to evolve to a specialized domain area, because the requirements will be high-volume, small-content and real-time asynchronous messaging. The technologies that are more capable of scaling to the value chain level or even to industry-level synchronization are likely to emerge from the transaction delivery network (TDN) and Web services network (WSN) providers.
Data Pooling, Cleansing and Normalization: These three steps are grouped because the technologies are combined when enterprises build Web-based catalogs.

Data pooling is the gathering of all necessary product data into one central location for a single enterprise (if large enough) or for multiple enterprises in a third-party solution, such as an e-marketplace or hosting service. (The necessary data is likely to be a subset of the total data associated with a given product.) This step involves the repetitive extraction and loading of heterogeneous data types of multiple formats from multiple systems into a common repository. Several technologies will continue to improve the efficiency of the process, although it will never be free of human labor. Other technologies that affect the maturity and evolution of this stage include artificial intelligence search capabilities, natural language recognition, image recognition, 3-D bar-coding, radio frequency identification (RFID), and the expansion of network, storage and computing capacity. However, all enterprises that decide to synchronize their product data through methods other than via proprietary methods and traditional point-to-point EDI will have to create formal data pools.

Cleansing and normalization involve the laborious process of homogenizing diverse data into a consistent structure with uniform units of measure. Circa 2015, the value of this function — from the cleansing of historic transaction data to supporting loads of (partial) trading partner and registry product descriptions — will create whole swatches of semi-automated offerings to parse pooled data and prepare it for registration. Cleansing and normalization will result in the final preparatory stage by which data is completed and made compliant to specific standards: Data gaps will be highlighted for filling in, and errors in data format and content will be highlighted. Exception processing becomes the most-valued feature, although manual labor remains to resolve errors. This will grow to support diversity in the sources and pools of product data within the enterprise and support propagation to internal applications and trading partners.

By 2013, original equipment manufacturer (OEM) extraction, transformation and loading (ETL) tools will be combined with native decision support and self-learning capabilities to dramatically reduce the manual labor required to load and validate item data. This will become a ubiquitous capability circa 2008. Publish and subscribe architectures will compress the period between product data loads from months (circa 2007) to hours (circa 2013) within the enterprise.

Large enterprises may need to establish these data pool environments internally to compensate for the heterogeneity born of divisional autonomy and application heterogeneity. In fact, the entire ecosystem will be made up of numerous certified data pools. There may be a different capability matrix for external industry registries.

Compliance and Endorsement: Compliance ensures that participant product data is in conformance with accepted industry standards. This is an independently verifiable stage. Compliance must also support individual enterprises’ quality assurance processes for inbound and outbound product data; however, the methods employed to do this will diversify, given the variability in relevant data volumes and business criticality. For strategic categories, this will be among the most-highly valued, as domain-specific applications persist in retaining control over portions of the product data life cycle. Over time, compliance will begin to exhibit bidirectional capabilities, as the range of PCDM capabilities extends to diverse data types that enrich product understanding. Requirements will be based on application type (for example, historical analysis vs. inbound logistics), diverse data types (such as RFID), manufacturing process definition and item maintenance histories.

Product Registries: Like a telephone directory, a registry acts as the main source for users and automated applications, enabling them to find products, determine the owners of products, and locate
where products are being bought and sold. At the highest industry level, the registry will be made up of multiple registries — perhaps distributed by country — all connected to a single, large virtual registry used to synchronize them all. On a small scale, a single enterprise may replicate this functionality within its PCDM application stack behind the firewall, so that its other departments and applications can ensure that all product data is kept synchronized around the enterprise.

Multiple registries will continue to develop as shared resources among trading groups, especially those based on common items within a particular industry. Although multiple industry or trade registries will be active, the distribution of product data and PCDM functions will be highly diverse. That is, there is simply too much product data in a given industry or trading community and not enough inherent benefit to justify the creation and maintenance of a single common product library. Thus, there will be diversity in the ways that registries are created, owned and managed. Diversity will be a product of variability in the volumes of data necessary to sustain synchronization, the business practices and cultures associated with given categories, variability in enterprise buy-sell strategies and trading-partner power positions.

In a few cases, some high level of cooperation among industry participants will create the need for distributed, local registries connected to one global registry — this is the case with CPG/retail and the efforts under way with UCCnet and driven by Global Commerce Initiative (GCI). In these projects, the physical, logical, performance and economic limits of post-modern computing infrastructures will prohibit the ubiquitous adoption of a highly centralized directory structure.

The era of treating product data as a departmental or enterprise matter will be drawing to a close, making way for the cross-enterprise adoption of PCDM. Users have been building data warehouses and deploying ERP as part of the concept of a single source of product data, as opposed to multienterprise, synchronization-dependent modes of product data management. Although internal management of product data is a nagging issue, users should begin to focus on and experiment with addressing the problem from a multienterprise perspective. Some industries have evolving models on which to achieve this and can serve as a model for understanding the future modes in which product data will be handled.

**Bottom Line:** Product content and data management (PCDM) is beginning to be recognized as a vexing and important multienterprise business-to-business challenge. The technology needed to manage product data inside and among enterprises is immature, but it will evolve and emerge in complementary stages — partly application-based and partly services-based. Enterprises should monitor the adoption of the PCDM technologies and services and begin building the skills necessary for selective adoption as they mature.