Development of Global Component Parts Libraries through Product Data Standardization

Presented by:

XSB, Inc.
21 Bennetts Road – Suite 100
Setauket, NY 11733
www.xsb.com

Introduction

XSB, Inc. (www.xsb.com) provides robust, scalable software that is critical for managing parts data used in the supply chain. Our customers often have independent systems for design, procurement, supply chain management, regulatory conformance and other product related activities. XSB tools help provide the semantic transformation necessary to make these systems interoperable.

The XSB solution captures, maintains and leverages a single, trusted, accurate, current and complete view of components found throughout the organization, its customers, and it's Web of suppliers. This is done by generating a global Master Data File. XSB defines Master Data as a consistent and standardized set of identifiers and extended attributes that describe the core entities of an enterprise (customers, products, specifications, suppliers, etc). Master data is particularly difficult to manage for enterprises that have grown through mergers and acquisitions. Acquired companies frequently employ different IT systems and post-acquisition enterprise data is frequently fragmented into incomplete, inconsistent, inaccurate and out of date, location-specific data silos. Lack of product data standardization between locations and divisions of companies have made it difficult for customers to realize the savings they expected from their Enterprise Resource Planning (ERP) implementations.

Developing a complete Master Data File is complicated when the 'correct' data resides outside of the organization at supplier or customer websites where other data inconsistencies may exist. Integrating this data is difficult because major search engines do not handle product semantics; they do not allow for the identification of unique products on the page, precise side-by-side comparison between products, or integration with internal enterprise product data. XSB technology rapidly aggregates and transforms internal and web data into uniform, accurate and semantically consistent data that is actionable across the enterprise.

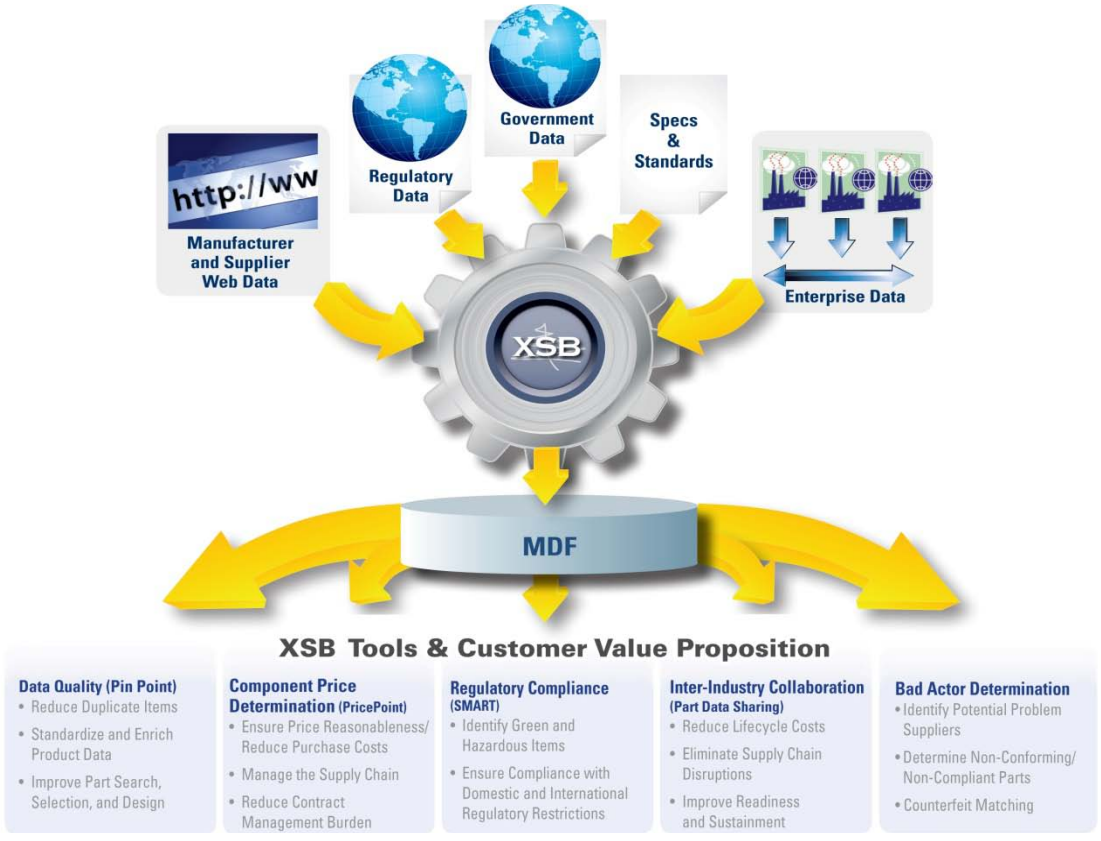

Figure 1: The XSB process of Master Data generation enables companies to optimize their part selection and price determination increase supply chain velocity while reducing supply chain interruptions caused by product obsolescence.

The XSB Technical Approach

The first step in Master Data File generation is the identification of procurable parts across the enterprise. Figure 2 below, shows a portion of the supply web for a TYCO connector. Note that different data sources describe the manufacturer for this connector as Tyco, TE Connectivity, or AMP (AMP is a brand of Tyco Electronics, and TE Connectivity is its current name). This is a common problem across the span of supplier names which has resulted in twenty percent duplicate parts in some enterprise systems. The supply chain for one XSB customer has over 300 different ways to spell Hewlett Packard (HewPac, HP, HP Printers…) and more than 50 ways to spell 3M Company. XSB standardization enables companies to identify and eliminate these duplicates.

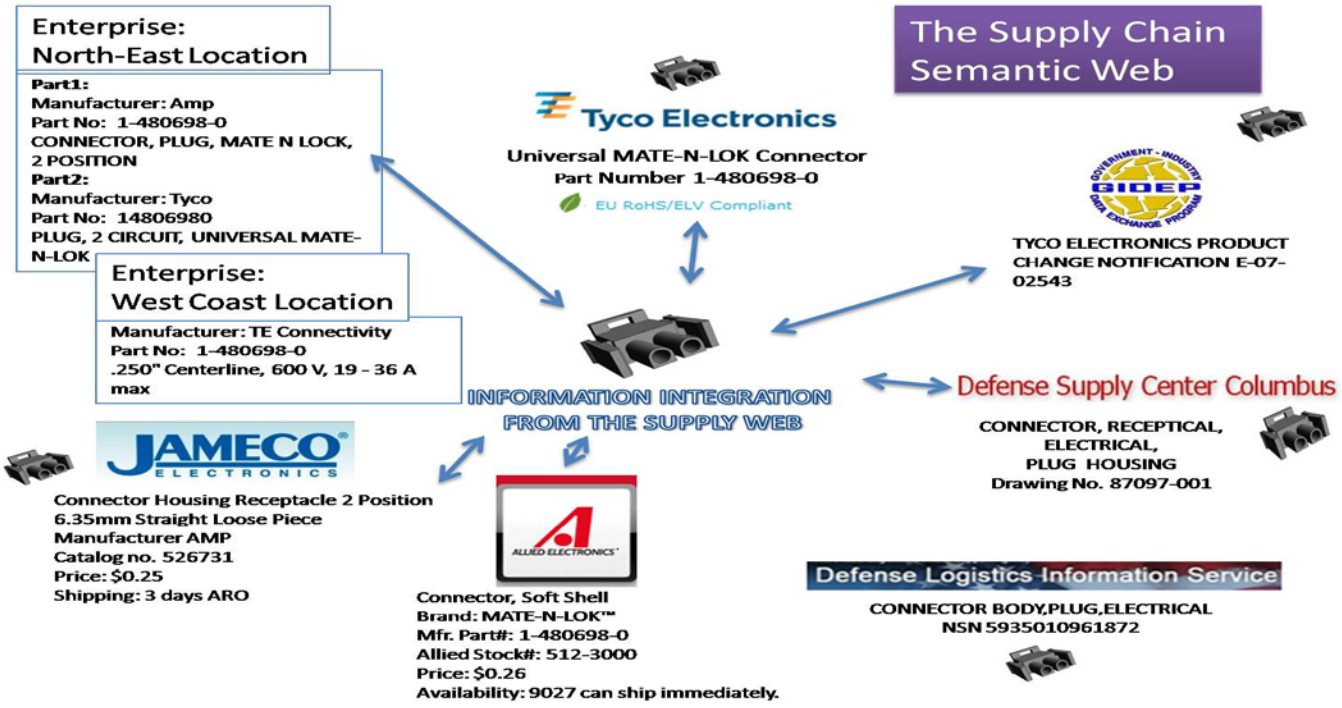

Figure 1

In addition to reducing duplicate parts by standardizing manufacturer names, brands, company associations and part numbers, the XSB System:

- Matches internal data to external data such as:
  - Data from Manufacturer and Supplier Web sites to provide improved characteristics and logistics information
  - Obsolescence and Counterfeit alerts for improved Product Lifecycle Management
  - Regulatory Restrictions for identification of Green and hazardous products including RoHS (Restriction of Hazardous Substances Directive) and REACH compliance
- Classifies descriptions to a customer specified ontology and standardizes context specific characteristics
- Generates an aggregated attribute view for every part in the Master Data File
- Determines reasonable prices for commodity products

The XSB process *of* ***Semantic Integration*** transforms disparate, heterogeneous data sources into coherent data according to a common ontology. Once the information is uniformly described it is possible to automatically analyze the product data, efficiently search for parts and reason about product interchangeability or the impact of regulations. XSB, Inc. has developed the patented ACT (Acquire, Classify, Transform) framework for this purpose (See Figure 3).

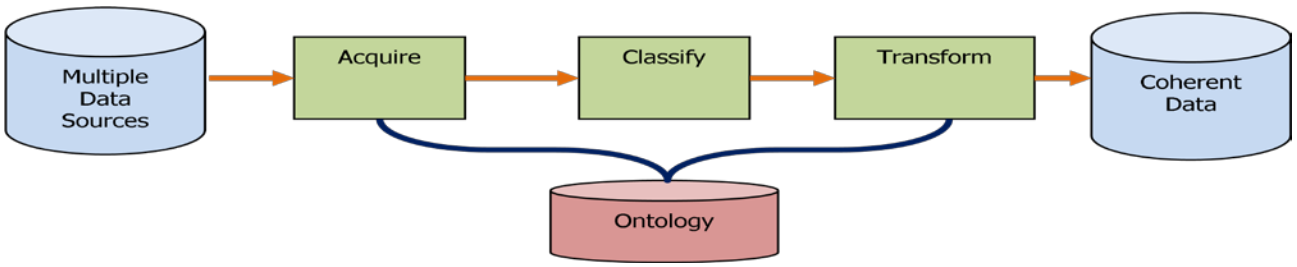

Figure 2: An ontology representing shared knowledge of the enterprise drives the ACT framework

The ***Acquire*** component in the framework captures the data (to be transformed) from structured, semi-structured, free-text, and Web data sources. The ***Classify*** component organizes the data into categories (classes) of similar products. Establishing the class of an object enables the system to determine the relevant attributes to be extracted. The ***Transform*** component extracts product attributes and standardizes them to a uniform representation. The result is a true ***Master Data File*** of de-duplicated, rich, standardized and easily searchable data. It is used for a wide range of purposes – from rationalizing the supply chain to the development of common parts libraries for engineering part selection. These components are described in more detail in the following section.

Data Acquisition:

The emergence of the WWW as the dominant medium for e-commerce has resulted in an explosion of independent on-line part data sources. The WWW has become a highly economical and efficient medium for posting part specifications, but few enterprise part management systems are well integrated with data from the Web. XSB has developed a proprietary ***Focused Crawler Technology*** for the automated acquisition of parts data available on the Internet. Unlike traditional approaches that crawl arbitrarily and treat web pages as documents, XSB Focused Crawlers are optimized to search only for product data. XSB Focused Crawlers automatically identify product pages and distinguish the relevant part content in each page. They exclude non-part information as well as information on unrelated parts. The information retrieved from a product page includes part identification (manufacturer name, part number), technical characteristics, prices, availability, and information about environmental compliance such as RoHS. Part information that is spread across more than one catalog page is automatically aggregated in the database. Part information for multiple parts on the same page is also recognized and correlated with different part numbers in the database. XSB Focused Crawlers do not require user input or complex navigation scripts to locate and extract product information. Focused Crawlers utilize graph-theoretical ideas to identify and cluster product pages within a catalog web site and then reason about web page structure to automatically identify and extract product data. Thus there is no need for manual encoding of site structure as is done by typical agents or crawlers.

Ontology Directed Classification:

The second critical step for generating coherent data is resolving the incompatibilities in structure and meaning of data from different sources. Independent data sources present similar information in different ways. This variety of data presentation for the same product becomes an impediment if a user wants to search by product characteristics or find similar parts.

The *XSB Ontology Directed Classifier (ODC)* groups similar items together by classifying item descriptions to a user specified taxonomy. For example, the ODC will organize all AC motors in the 'MOTOR, ALTERNATING CURRENT' category or all hex head screws into a category of 'SCREW, HEXAGONAL HEAD'. The patented XSB ODC uses an algorithm based on statistical analysis of word frequencies in taxonomy labels and associated training data. This algorithm improves on naïve Bayesian classifiers by utilizing word placement in the taxonomic structure. ODC classification is tunable by providing common word abbreviations and correctly classified items as training examples. The system is fast, scalable, and easily handles taxonomies of tens of thousands of nodes, with up to a hundred thousand training items. Once items from a diversity of sources have been classified, their descriptions are used to extract properties or attributes of these items based on where the product is classified.

Ontology Directed Extraction:

The *XSB Ontology Directed Extractor (ODE)* is used to discover attributes in free text descriptions and standardize them to a common schema guided by the Ontology. Extraction is based on the information about a class in the ontology to which the objects belong. The ontology stores information about classes or categories of objects for which extraction is to be done. It describes the properties that are appropriate for each class of objects and the types of acceptable values for each property. In addition, the ontology contains information about abbreviations and special word usages, the contexts of their use, and disambiguation information. ODE uses this information to recognize attributes in part descriptions and to extract their values.

Values for the same attribute may be described in a variety of ways using abbreviations, synonyms, and differing conventions. For example, two screw descriptions could be stated as ***.025 dia x 1-/14 long, 20 tpi*** and ***¼-20 x 1.25".*** ODE transforms these descriptions into structured, consistent data:

ATTRIBUTE NAME	VALUE
NOMINAL THREAD DIAMETER	0.250 INCHES
THREAD QUANTITY PER INCH	20
FASTENER LENGTH	1.250 INCHES NOMINAL

Conclusion

The XSB system is a robust, cloud based ontology directed system for managing complex supply chains. The system scales to ontologies with tens of thousands of nodes and operates on part libraries of nearly one hundred million component parts.

The XSB solution captures, maintains and leverages a single, trusted, accurate, current and complete view of components found throughout the organization, its customers and its Web of suppliers by generating a semantically consistent Master Data File. Once parts information is uniformly described in the MDF, it is possible to automatically analyze product data, efficiently search for parts and reason about product interchangeability or the impact of regulations. The result is a significant improvement in product selection, design, pricing, inventory, and obsolescence management.

For More Information:

Daria Wesnofske
631-371-8102
d.wesnofske@xsb.com

Len Bosma
631-626-2242
l.bosma@xsb.com