

CASE STUDY

Mastering the Challenge of Product Data Consolidation

The Goal: Streamline procurement through consolidating the product inventory

The Task: Consolidate a product inventory of over 100,000 stock keeping units (SKUs), stored in four different databases, formats, and languages

The Challenge: Create a new taxonomy; aggregate, clean and standardize the data; classify it into categories; extract product attributes; enrich the data; identify duplicate or similar products; and create consolidated, tabulated product data

The Team: InQuera Data Masters: Specialists in product data quality, data governance, and taxonomy

The Tools: The DataRefiner Suite: High-performance Automated Product Data Quality System that transforms dirty data into high-quality master data

The Results: One new taxonomy with a hierarchy of categories, attributes, and attribute values; comprehensive, consistent, normalized descriptions; products were classified according to new taxonomy; and 10% of duplicates identified

Iscar, the second largest manufacturer of precise carbide metal cutting tools, had a product inventory of over 100,000 SKUs. With four manufacturing plants located in Israel, Germany, Korea, and the United States producing similar products using similar technology, the likelihood of duplication was very high.

But the product descriptions, stored separately in different sites under different classification systems, were unclear, incomplete, and inconsistent, making it impossible to find or compare identical or even similar products. Furthermore, the Israeli data contained partial attributes, attribute values, and free text descriptions, while the data from other countries only contained inconsistent free text descriptions.

How was Iscar to tackle the task of consolidating its product — especially when the information was stored in separate databases, diverse formats, and different languages? Consolidating the inventory — 60,000 in Israel and 40,000 from the German, Korean, and U.S. plants — was a mammoth task and could not be handled in-house. Iscar contacted InQuera, data masters with specialized knowledge in the MRO market. InQuera's Product Data Quality Suite, DataRefiner, met Iscar's need for a comprehensive data refinement process that combined the four databases into a single source, with organized, classified, and standardized product data.

According to Assaf Turtel, Director of Content, InQuera, the first task was to take Iscar Israel's existing taxonomy and review it. The Israeli database, written in both English and Hebrew, had a taxonomy with partial attributes and values. (A taxonomy is a classification scheme that groups similar items into hierarchical categories according to established criteria and assigns relevant parameters to each category.) The other plants had basic category hierarchies (different at each plant) and free text descriptions but no attributes or attribute values.

InQuera's analysis showed that the taxonomy did not enable a clear enough classification of items. A new, normalized taxonomy using the DataRefiner Taxonomy Manager was developed, which proved to be a very challenging task, none the least because the Iscar team had grown used to their taxonomy and had to adapt to a new structure. An important aspect of the new taxonomy was the definition of permissible attributes for each category and their respective attribute values. While the original Iscar taxonomy classified product according to usage, the new taxonomy classified the product itself, *irrespective* of usage.

In the next phase of the project, the team reclassified the 60,000 items in the Israeli database to the new taxonomy using the DataRefiner Classifier. The Classifier, an active learning system utilizing artificial intelligence (AI), leveraged knowledge gained from the Israeli database to auto-classify data contained in the other databases. Once Iscar authorized the reclassification, the team reclassified the 40,000 items from the other companies and merged the items into one comprehensive repository under one taxonomy hierarchy.

At this point, Iscar reviewed its inventory to understand what categories represented the highest spend for the company. This analysis resulted in reducing the inventory to be fully refined from 100,000 to 60,000 items.

Once the product data had been reclassified, the team mapped the existing attributes and attribute values contained in the Israeli data to the repository and normalized the attribute values using DataRefiner Formalizer, before extracting more information from the description to enrich the attributes.

What is data refinement?

InQuera's data refinement process aggregates product data from different languages, sources, applications, and formats; cleans and standardizes the data; classifies it into categories according to the taxonomy (existing or new); extracts product attributes; enriches the data; and identifies duplicate or similar products, to create consolidated, tabulated product data.

Once that step was accomplished, the team extracted the attributes and attribute values from the free text descriptions of the other plants using DataRefiner Extractor, ensuring that the attribute values adhered to the approved content structure. The Extractor, an AI-based system, enabled the domain experts to define context-oriented rules to automatically extract pieces of information from the free text descriptions into pre-defined parameters.

Once the refinement process was completed, the result was an organized table with attributes and attribute values for all the leaf (lowest level) categories. Deduping, the process of identifying duplications, was the next product of the process. To prepare for deduping using DataRefiner Deduper, the team had to develop a strategy to define what product are the same or similar by assigning a key identifier such as the manufacturer or supplier number. In addition, key parameters had to be defined per category as well as across the entire repository (through primary attributes). The challenge was to identify the key parameters and ensure that there was sufficient important information available to enable the comparison and make a decision. (If important information is only available for a small number of items, there is no point in making a comparison.)

InQuera provided Iscar with deduping tools so that they would be able to handle deduping in-house, based on the InQuera strategy. Some 10% of duplicates were identified. Moshe Bram, Technical Support Manager for the Iscar Purchasing Center, was greatly impressed with InQuera's capabilities. All products now have comprehensive, consistent, normalized descriptions and are classified according to the new taxonomy. From four separate databases, Iscar now has centralized, refined data in an easy to use format.