

WHITE PAPER

THE Dirty Data Trap

Tools cannot fix what process, standards,
and ownership have not touched.

Product Data: What It Is and Why It Is Messy

Product data tells the world what an item is and its attributes:

- Its size, material, and technical specs
- The standards it must meet
- How it is packaged, shipped, and shown online

When this data is messy, everything slows down. Onboarding, search, compliance, and even AI stumbles. When it is clean, products move smoothly from supplier to shelf to checkout.

The trap is simple. Many companies buy new systems expecting product data problems to go away. But the same issues return, because the data itself never changed and governed.

- Search fails when “red” was never entered as an attribute.
- Filters break when “10in” and “10 in” show up as separate options.
- Products sit in the wrong category and inherit nothing.

New tools don't fix the root cause of the problem.
And AI doesn't either. It only magnifies the quality of whatever you feed it.

The Cost of Messy Product Data

Messy product data creates pain across the business:

- Missing attributes keep products out of search.
- Inconsistent values create duplicate filters.
- Misclassified products inherit the wrong attributes.
- Wrong details increase returns.
- Teams waste hours patching gaps by hand.
- Sales and customer service need to provide missing information to customers.

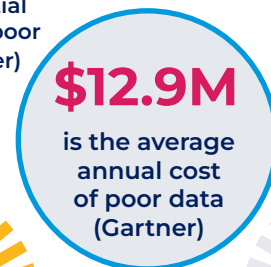
THE NUMBERS SHOW THE IMPACT:



less revenue potential for companies with poor taxonomy (Forrester)



improvement in search results after optimizing schema (Gartner)



is the average annual cost of poor data (Gartner)



Poor data quality costs 20-30% of revenue each year (Experian)



fewer customer support calls after improving product data

Lessons From the Frontlines

We've seen firsthand how messy product data derails sales.

Take the case of the missing word "red."

We tested a distributor's site the way a customer would. We searched for a red coaxial cable. Even though they had both six-inch and six-foot versions in stock, nothing appeared. Why? The "red" attribute wasn't filled in.

We ran the same search on a competitor's site. Their data was complete, the "red" attribute was in place, and the cable showed up instantly. The competitor won the sale, while our customer's inventory stayed invisible.

We've also seen:

Scattered Attributes

Some companies list product attributes alphabetically on PDPs instead of grouping them. That splits related specs apart—"cranking amps," "marine cranking amps," and "cold cranking amps" land in different places on the page. Without grouping, it's harder for buyers to scan and compare values.

Funky Filters

One site had five different versions of "aluminum" in a single filter set—misspellings, abbreviations, stray spaces. When filters only allow single-select, buyers can't see all relevant products.

Missing Attributes

In a review for an HVAC manufacturer, furnaces were missing heating capacity, airflow, and dimensions. Humidifiers lacked output, coverage, and size. A customer landing on those PDPs had almost nothing to go on, they'd need to call support. Most would leave.

Vague Values

On a grocery retail site, we found filter menus stuffed with "Other." In the eggs category, "Other" sat right alongside "Organic" and "Cage Free." Shoppers had no idea what "Other" meant, so the filter did nothing to guide their choice—and many clicked away.

Messy product data surfaces right where customers interact with it. It breaks the buying experience and costs sales.

Why Cleaning Alone Doesn't Fix Product Data

Correcting typos or merging duplicates buys time, but it doesn't solve the root problem. New systems simply carry the mess forward unless data is built on a solid foundation.

Strong product data requires:

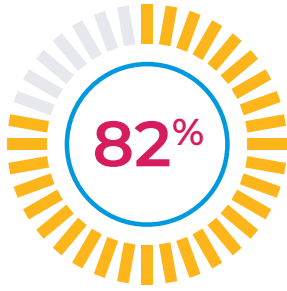
- A proper taxonomy and schema to structure attributes consistently
- Enrichment with category-specific attributes fully captured
- Completeness with no missing values in critical fields
- Governance with clear ownership and accountability
- Delivery that is consistent across all channels

And getting there is not a one-time task. It takes ongoing enrichment, governance, and regular audits. Automation can help, but it cannot replace the foundation.

Clean product data requires discipline and repetition, not just another tool.

Making the Work Manageable

Cleaning and governing product data is hard work. Research shows:



82% of leaders say poor ownership slows decisions (Experian)



Data quality issues consume 40–60% of data team time (Gartner)



Up to 50% of IT resources are wasted managing “data debt” (IDC)

The fix requires people and automation working together.

People set the rules. They decide what attributes should look like, how deep the taxonomy should go, and how to handle edge cases. For example, grouping “marine cranking amps” with “cold cranking amps” takes human judgment.

Automation enforces the rules. It standardizes units, removes spelling variations, applies formatting, and flags errors. It can stop “10in” and “10 in” from ever becoming separate filter values.

People build the foundation. Automation keeps it clean at scale. Together, they prevent old problems from coming back.

A 30-60-90 Day Plan



Assess and Map

- Audit product data in top categories
- Document all systems that feed or use attributes
- Identify the ten most critical gaps costing sales or slowing onboarding



Fix the Foundation

- Redesign schemas using the 80/20 rule (if 80% of products need an attribute, include it)
- Separate product and packaging dimensions with clear field names
- Build a new item onboarding workflow with stage gates and accountability
- Write style and spelling rules to prevent messy filters



Scale and Show Value

- Launch grouped attributes on PDPs so buyers can compare specs easily
- Apply industry standards such as automotive fitment, electrical codes, and grocery packaging
- Pilot auto-classification with feedback loops
- Publish a dashboard showing gains in onboarding speed, return rates, and search-driven conversions

The ROI of Clean Product Data

The benefits are proven:

- ✓ **15% higher order value** from better taxonomy and filters (BigCommerce)
- ✓ **42% of frustrations** come from missing product info (DHL)

NAED \$5B Impact Report:

- ✓ 32% faster adoption of new products
- ✓ 32% increase in order size
- ✓ 23% more reorders
- ✓ 21% more repeat purchases
- ✓ 10% lift in related product sales

Pivotree benchmarks:

- ✓ 60% faster onboarding
- ✓ 20% more search-driven conversions
- ✓ 30% fewer returns
- ✓ 15% more organic traffic
- ✓ 99.7% accuracy with governance in place



Where SKU Build Fits

SKU Build combines human expertise with automation. Experts define the rules, attributes, and structures. Automation enforces them at scale.

Works with your existing systems

Fixes the foundation and keeps it clean

Stops the cycle of bad product data moving from one system to the next

IN SHORT:

SKU Build makes product data clean, structured, and ready for search, filters, AI, and every system downstream.



About Pivotree

Pivotree, a leader in frictionless commerce, strategy, design, build, and managing digital Commerce, Data Management, and Supply Chain solutions for over 150 major retailers and branded manufacturers globally. With a portfolio of digital products as well as managed and professional services, Pivotree provides businesses of all sizes with true end-to-end solutions. Headquartered in Toronto, Canada, with offices and customers in the Americas, EMEA, and APAC, Pivotree is widely recognized for its partnership with top brands across industries.

For more information, visit www.pivotree.com or follow us on [LinkedIn](#).



sales@pivotree.com | 1-877-767-5577 | www.pivotree.com

©Pivotree Inc. All rights reserved.