Product Data Quality - Different Problem, Different Solutions

ABSTRACT

In discussions of Data Quality, it is often assumed that the same tools and techniques that work well in one data domain will work well in any data domain. Specifically, it is often assumed (and sometimes asserted) that experience dealing with customer data is a qualification to also deal with product data. In practice, this is hardly ever the case - as reflected in the emergence of domain-specific markets, toolsets and techniques. Through a series of real-world customer case studies and experiences, this session will explore the differences and similarities between product data and other data types in terms of the core data quality problems, use cases, business drivers and benefits as well as the available tools, techniques, architectures and best practices available to address them.

BIOGRAPHY

Martin Boyd
Vice President of Marketing
Silver Creek Systems

Martin Boyd leads strategic positioning and market expansion for Silver Creek Systems, a technology leader in product data quality. His fifteen plus years in Product Management and Strategic Marketing for enterprise software companies brings broad understanding of the practical data issues and real-world solutions required to standardize and integrate disparate data sources within the information supply chain.

Prior to joining Silver Creek Systems, Boyd successfully led the repositioning of Ariba, a spend management technology company, and has held several senior and management marketing roles for companies such as Oracle, Lucas Management Systems and IBM. Martin holds a Masters degree in Engineering from Strathclyde University.
Product Data Quality
- Different Problem, Different Solutions

Martin Boyd, VP Marketing, Silver Creek Systems

MIT Information Quality Symposium, July 15-17 2009

Agenda

Product Data Quality – A large and growing market

Product Data is Different
- Understanding the Product Data Problem
- Comparison with ‘traditional’ customer Data Quality

Product Data Quality – Applied Use Cases
- Leading Healthcare Alliance
- World-class Retailer
- Semantic Recognition & Transformation Examples

Next-generation Solutions for Product Data Quality
- Solution Requirements
  - Semantic understanding
  - Rapid learning
  - Integrated Governance
- Solution Demonstration
### MDM Segment Size & Growth (Gartner, Nov 08)

#### Customer Data Quality

<table>
<thead>
<tr>
<th>Segment</th>
<th>2007</th>
<th>5 Year CAGR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer (CDI)</td>
<td>$335m</td>
<td>21%</td>
</tr>
<tr>
<td>Product (PIM)</td>
<td>$401m</td>
<td>22%</td>
</tr>
<tr>
<td>Procurement</td>
<td>$36m</td>
<td>29%</td>
</tr>
<tr>
<td>Asset</td>
<td>$56m</td>
<td>28%</td>
</tr>
</tbody>
</table>

#### Product Data Quality

---

**Product Data is Different**

- Understanding the Product Data Problem
- Comparison with ‘traditional’ Customer Data Quality
Product Data Quality

- "One of the most difficult type of data to master is undoubtedly product data – the items, assemblies, parts and SKUs that are core to many businesses.
- Product data is inherently variable, and its lack of structure is generally too much for traditional, pattern-based data quality approaches.
- Product and item data requires a semantic-based approach that can quickly adapt and ‘learn’ the nuances of each new product category. With this as a foundation, standardization, validation, matching and repurposing are possible. Without it, the task can be overwhelming and is likely to include lots of manual effort, lots of custom code – and a whole lot of frustration."

Andrew White, Research VP, MDM and Supply-chain

The Product Data Problem

What is this?

10hp motor 115V Yoke mount
MOT-10,115V, 48YZ,YOKE
mtr, ac(115) 10 horsepower 115volts
This 10hp yoke mounted motor is rated for 115V with a 5 year warranty
10 Caballos, Motor, 115 Voltios
TEAO HP = 10.0 1725RPM 115V 48YZ YOKE MTR
Motor, TEAO, 1725 RPM, 48YZ, 15 Voltios, Montaje de Yugo, hp = 10

How many different ways can you receive information?
How many ways do you need it?
Typical Product Data Problems

The greatest threat to your PIM/MDM initiative

- Inconsistent names (representing same business)
- (Often) Rich information, but mostly non-standard
- Attributes non-standard, missing or invalid

<table>
<thead>
<tr>
<th>Product ID</th>
<th>Manufacturer</th>
<th>Description</th>
<th>Product Type</th>
<th>Power</th>
<th>Voltage</th>
<th>Mount</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC123</td>
<td>AA Inc.</td>
<td>10hp motor 115V Yoke mount</td>
<td>Motor, AC/DC</td>
<td>10hp</td>
<td>115V</td>
<td>Yoke</td>
</tr>
<tr>
<td>abc-123</td>
<td>A.A.</td>
<td>110 HP, horse power 115volts</td>
<td>DC Motor</td>
<td>10hp</td>
<td>115V</td>
<td>Yoke</td>
</tr>
<tr>
<td>ABC/123/Q</td>
<td>AA/Craft</td>
<td>TEAO HP = 10.0 1725RPM 115V 48YZ YOKE MTR</td>
<td>AC/DC Motor</td>
<td>10 HP</td>
<td>115V</td>
<td>N/A</td>
</tr>
<tr>
<td>Z99</td>
<td>Z99</td>
<td>M0T-10, 115V, 48YZ YOKE</td>
<td>Motor</td>
<td>Z99</td>
<td>115V</td>
<td>Yoke</td>
</tr>
</tbody>
</table>

- Inconsistent formats (extra characters often added)
- Misclassified item (not a motor)
- Inconsistent classifications & misclassifications

Widespread duplication (often hard to spot)

Companies struggle with the basics of PIM
- 80% companies are not confident in the quality of their product data
- 73% find it “difficult” or “impractical” to standardize product data

Product Data – Common Characteristics

- Multi-domain, Multi-attribute
  - 5 to 5 thousand “domains” per system
  - 2 to 30 required attributes per domain/item
  - Many requirements, validations, rules per attribute

  Too complex for custom code

- “Poor” data
  - Little structure
  - Few standards
  - Incomplete information

  Too ’messy’ for traditional tools

- High volumes/Turnover
  - 10k to 10M items
  - 5 to 500% turnover (annual)

  Too much volume for manual effort
Typical Product Data ‘Solutions’

The value of having ‘the right tool for the job’

Current methods [for product data quality] often don’t work
66% companies use "manual effort" or "custom code"
- 75% say it is too unreliable
- 64% say it is too slow
- 56% say it is too expensive
- >50% say ‘all of the above’

Semantic-based Product Data Mastering (DataLens™)
outperforms
- Manual Effort: Premier Healthcare is saving $4-5M/year in manual cleansing and matching
- Custom Code: Avnet replaced 5 years worth of custom code within 90 days and improved their match rate to quote on $millions more items
- Traditional DQ: A leading hospital group replaced a well-known DQ platform to avoid the constant, expensive and time-consuming script maintenance

Customer Data Quality vs. Product Data Quality

Different problems, different technologies

Customer Product (part, SKU, item, asset etc.)

- Name & Address Data
  - Relatively fixed syntax
  - Mis-spellings and name equivalents
  - Match based on probabilities

- Pattern-based tools work well

Product Data
- No fixed syntax – few standards
- Infinite variability – format, content, syntax
- Different rules for each product category

Must have semantic understanding and ability to learn new domains quickly

Silver Creek Systems ©2009
TDWI – Product Data Quality

- Product data differs from other domains, so it has unique uses and requirements
- Customer-oriented data quality techniques and tools can be retrofitted to operate on other data domains, but with limited success
- Standard data quality techniques don’t work with product data without significant adaptation
- Customer data has a base standardization, product data doesn’t
- Customer data has predictable patterns and syntax, whereas product data doesn’t.
- With data quality solutions, one size rarely fits all.
- Product data cleansing and standardization are greatly facilitated by a semantic approach
- Semantics-driven data quality capabilities are crucial with product data
- With product data, exception management is more complex and manual than with most data domains
- Product data transformations depend on the context of each process or application for which it is repurposed

MDM Implementation Effort

10% - MDM software implementation

40% - Establish governance & document master data architecture

50% - Data remediation

Clean-up to meet the new rules
- Find duplicates
- Eliminate discrepancies
- Fill gaps
- ...

Data Mastering

Getting your data right and keeping it right
Product Data Mastering - Semantic Architecture at the Core

Semantic recognition for extraction, standardization & transformation based on meaning and context

Get your data right and keep it right

Product Data Mastering

Quality Governance - Data Steward Dashboard

Exception Management - Validation & Remediation

Standardize
- Std. Classification(s)
- Std. Attributes
- Std. Descriptions
- Translate (inbound)

Match
- De-duplicate
- Merge

Enrich
- Classify
- Extract/append info

Repurpose
- Translate (outbound)
- (Re-)classify
- (Re-)standardize
- (Re-)format

Semantic Recognition - Contextual understanding & learning

Combination of Data Quality, Data Integration & Data Governance to recognize, cleanse, match, govern, validate, correct and repurpose product data from any source to any target

Silver Creek Systems ©2009

Product Data Quality - Applied Use Cases

- Leading Healthcare Alliance
- World-class Retailer

www.silvercreeksystems.com
Leading Healthcare Alliance – Premier Inc.

Excerpt of Presentation at Gartner MDM Conference, November 2009

Product MDM – The Practical Challenges

• **Lots of data**
  - Over 7 million items/SKUs
    - Each with up to 50 part numbers and an infinite number of descriptions

• **Many disparate data sources:**
  - 3 Legacy systems
  - 1200+ Suppliers
  - 50+ Distributors
  - 1700+ Hospitals
  - 40,000+ Healthcare Facilities
  - 10+ Third Party Data Suppliers

• **No Standards**
  - Supplier Identification Number
  - Product Identification Number
  - Packaging
  - Descriptions
  - Attributes…
Data Challenges – How Many Ways Can You Say “3M”?

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Identification Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allegiance</td>
<td>M8630</td>
</tr>
<tr>
<td>Owens &amp; Minor</td>
<td>4509008630</td>
</tr>
<tr>
<td>BBMC-Colonial</td>
<td>0450908630</td>
</tr>
<tr>
<td>BBMC-Durr</td>
<td>081048</td>
</tr>
<tr>
<td>Kreisers</td>
<td>MINN8630</td>
</tr>
<tr>
<td>Midwest</td>
<td>TM-8630</td>
</tr>
<tr>
<td>Pacific</td>
<td>3/M8630</td>
</tr>
<tr>
<td>UnitedUMS</td>
<td>001880</td>
</tr>
</tbody>
</table>

Different distributors and hospitals use different identification numbers

3M™ DuraPrep™ Surgical Solution [Iodine Povacrylex [0.7% available Iodine] and Isopropyl Alcohol, 74% w/w) Patient Preoperative Skin Preparation
**What is Data Mastering?**

Data Mastering: The ability to understand, recognize and 'handle' data from any source as a real-time process

- Standardize
- Enrich
- Match
- Classify
- Remediate
- Translate
- Validate
- Re-purpose
- Govern

**ID and Semantic Matching**

<table>
<thead>
<tr>
<th>Description</th>
<th>From the PO</th>
<th>Standardized to</th>
<th>Matched to</th>
</tr>
</thead>
<tbody>
<tr>
<td>REAGENT METHOTREXATE II</td>
<td>07A12-40</td>
<td>ABBOTT LABORATORIES, INC. 07A1260 EA</td>
<td>REAGENT TDM METHOTREXATE II FPIA ABBOTT TDX/TDXFLX 100TEST ABBOTT LABORATORIES, INC. 07A1260 EA</td>
</tr>
<tr>
<td>NEEDLE 15-DEGREE STAMEY</td>
<td>095015 ANNUAL</td>
<td>COOK GROUP, INC. 056015 EA</td>
<td>NEEDLE UROLOGY STAMEY 15DEG 6 3/4IN COOK UROLOGICAL INCORPORATED</td>
</tr>
<tr>
<td>NDL ANASTOMIC DVC 8PK</td>
<td>S20SFRN8 BX</td>
<td>MEDTRONIC USA, INC.</td>
<td>DEVICE ANASTOMOSIS TITANIUM CLIP SINGLE ARM SINGLE NEEDLE 20FR ID SHORT FLEX REGULAR NEEDLE 8/PK MEDTRONIC USA, INC. S20SFRN8 PK</td>
</tr>
</tbody>
</table>
Semantic Match with Alternatives

Data Mastering for Premier
Semantic-based Data Mastering
(using Silver Creek Systems)

- **Enhanced approach** solves many persistent problems
  - Semantic-based approach handles unstructured & non-standard data
  - Rapid deployment across hundreds of categories
  - Specialized matching capabilities eliminates significant manual effort

- **Ease of use** puts business rules in hands of business user
  - Clinical experts can maintain rules
  - Less dependence on IT

- **Fast deployment** for early ROI
  - ‘Auto-learn’ capability reverse-engineers rules from existing data
  - Phased deployment category by category

- **Flexibility**
  - Make changes in minutes instead of months
  - Services-based deployment can plug-in to any process

World-class Retailer

www.silvercreeksystems.com
World-class Retailer

Problems:
- Review of website show significant ‘blind spots’ (insufficient product attributes for guided navigation)
- 20 weeks to put merchandising rule changes into production
- Need to rapidly scale-up website SKU-count

<table>
<thead>
<tr>
<th>SKU range</th>
<th>Dimensions</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Grand Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;25</td>
<td>297</td>
<td>16</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>416</td>
</tr>
<tr>
<td>25-49</td>
<td>188</td>
<td>29</td>
<td>15</td>
<td>7</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>246</td>
</tr>
<tr>
<td>50-74</td>
<td>41</td>
<td>6</td>
<td>9</td>
<td>8</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td>75</td>
</tr>
<tr>
<td>75-99</td>
<td>25</td>
<td>14</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td></td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>48</td>
</tr>
<tr>
<td>&gt;100</td>
<td>12</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>26</td>
</tr>
<tr>
<td>Grand Total</td>
<td>527</td>
<td>89</td>
<td>48</td>
<td>80</td>
<td>29</td>
<td>12</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td></td>
<td></td>
<td>811</td>
</tr>
</tbody>
</table>

Blindspots 230 24 2 4

- 65% of categories have no category-specific navigation
- Half of these categories have > 25 SKUs

The Challenge: Category-specific Information

Universal Navigation
- Apply to all items across all categories
- Simple to generate and deploy

Category-Specific Navigation
- 2-10 dimensions across thousands of categories
- Thousands of mappings
- Tens of thousands of validation rules
- Millions of recognition rules
- Complex workflow to resolve missing or incorrect information

Universal Dimensions

Category-specific Dimensions

BUT Rarely adequate for “World class” guided navigation

Much greater complexity, BUT huge payoff in guided navigation and customer experience

Silver Creek Systems ©2009
Category-Specific Information

Strap Type

Closure Type

Size

Material

Color

Character

Drop Length

Colors
- apple
- foamy aqua
- asparagus
- Bronze
- cement
- Cognac
- eggplant
- emerald
- Expresso
- faded yellow
- fuchsia
- Ivory
- jade
- khak
- lavender
- bright lemon

Standardized Colors
- Green
- Blue
- Green
- Brown
- Gray
- Brown
- Purple
- Green
- Brown
- Purple
- Yellow
- Pink
- White

Attributes must be extracted and standardized

Sample Standardizations

Drop Length
- 9" drop
- 9 to 11 inches
- 12" drop
- 13" drop
- 19" drop
- 20" drop

Standardized Drop Length
- Under 6 inches
- 6 to 8 inches
- 9 to 11 inches
- 12 inches or More
Live with 1,000 product categories in 10 weeks

Extract + Standardize + Maintain + Map

- Significant improvement in data quality & 'searchability'
- Huge savings in manual data preparation
- Scalable process & infrastructure
- 40x faster merchandising rule changes (20 weeks to <24 hours)

Quick Stats:
- > 1,000 product categories covered
- Extracting 1684 attributes across categories
- 65,000 SKU's processed in 45 mins
- 10 weeks from training to go-live

Semantic Recognition & Transformation Examples
Step 1 – Semantic Recognition of Category

1. Semantic recognition uses whole context to determine meaning and category

<table>
<thead>
<tr>
<th>Input Description</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>4&quot; end cap</td>
<td>Piping &amp; plumbing</td>
</tr>
<tr>
<td>4 in cap screw</td>
<td>Fasteners</td>
</tr>
<tr>
<td>4 in box b-ball cap</td>
<td>Headgear</td>
</tr>
<tr>
<td>750ml Merlot, screw cap</td>
<td>Capacitors</td>
</tr>
<tr>
<td>750ml Merlot, screw cap</td>
<td>Wine</td>
</tr>
</tbody>
</table>

2. Information extraction uses category-specific semantic models to identify and extract target information

Extracted Attributes
- Diameter = 4 inch
- Length = 4 inch
- Quantity = 4
- Capacitance = 4ESU = 4.45 picofarad
- Volume = 750ml, varietal = Merlot

Missing information
- Material; Thread
- Thread; Diameter; Material
- Color
- Manufacturer; Mount type; Operating characteristics

3. Category-specific semantic models flag missing information for potential remediation

Standardized Description
- End cap, 4 inch
- Hex cap screw, 4 inch
- Baseball cap, 4 per box
- Capacitor, 4.45pf (4 ESU)
- Wine: Merlot, 750ml, screw cap

4. Data is converted, transformed and re-assembled according to requirements of target system

Step 2 – Semantic Transformation to Target Format

Sample: Resistors

<table>
<thead>
<tr>
<th>ITEM DESCRIPTION</th>
<th>Key attributes extracted and normalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resistors 220 Ohm 6.25 Watt Axial Carbon Film</td>
<td></td>
</tr>
<tr>
<td>Resistors 220 Ohm 6.25 Watt Axial Carbon Film</td>
<td></td>
</tr>
<tr>
<td>Resistors 50 Ohm 1% 1/4 Watt Film</td>
<td>Key attributes extracted and normalized</td>
</tr>
<tr>
<td>Resistors 50 Ohm 1% 1/4 Watt Film</td>
<td>Classified to many number of schemas – industry standard or custom</td>
</tr>
<tr>
<td>Resistors 50 Ohm 0.1% 1/4 Watt Axial Carbon Film</td>
<td>Normalized descriptions to fit any standard or format</td>
</tr>
<tr>
<td>Resistors 220 Ohm 25 Pk Network</td>
<td>Translated into any language (incl. double-byte languages)</td>
</tr>
</tbody>
</table>

Semantic recognition identifies item category and key attributes even with highly variable data

Item can then be cleansed (standardized) as desired
Step 2 – Semantic Transformation to Target Format

**Original Description**

```
TEAO, 1HP, 1725 RPM, 115V, Yoke, NPT 1-1/4" LH, 1-1/4" RH, 3/4" CH, 1-1/2" WH, TKEO
```

**Enclave**

```
Encl. Name  Power  Speed  Frame Size  Voltage  Frequency  Bearing
TEAO  1440  1725 RPM  40HP  1 15 V  Yoke
TEAO  1440  1725 RPM  40HP  2 15 V  Yoke
TEAO  1440  1725 RPM  40HP  3 15 V  Yoke
TEAO  1440  1725 RPM  40HP  4 15 V  Yoke
TEAO  1440  1725 RPM  40HP  5 15 V  Yoke
```

**Key attributes extracted and normalized**

**Sample: Motors**

```
Highly variable syntax (patterns)
Semantic recognition identifies item category and key attributes even with highly variable data
Item can then be cleansed (standardized) as desired
```

**Normalized descriptions to fit any standard or format**

**Sample Description**

```
```

**Translated into any language (incl. double-byte languages)**

**Step 2 – Semantic Transformation to Target Format**

**Product Description**

```
Flexible Acetate 5-Ring-Ring Binder, 1" Capacity: 8-1/2" x 11" Black
Flexible Acetate 5-Hole Ring Binder, 1" Capacity: 8-1/2" x 11" Black
Flexible Acetate 5-Hole Ring Binder, 1" Capacity: 8-1/2" x 11" Black
Flexible Acetate 5-Hole Ring Binder, 1" Capacity: 8-1/2" x 11" Black
Acetate Handle + Channel: 42-7/8" of Double Hanging Data Binder, 1" Black
Metal-5-Spring Label Holders for 15" and 1" Binders: Clear Vinyl, 12-Pack
Metal-5-Spring Label Holders for 7-1/4" and 2" Binders: Clear Vinyl, 12-Pack
Wide-220 Ring Presentation View Binder, 1" Capacity: Letter, White
Double-Fastener Centers for活Binding, 8-1/2" x 11" Black, 5 Per Set
```

**Enclave**

```
Type  Ring Type  Rings  Sheet Capacity  Sheet Material  Lock  Inside Label  Color  Packaging
Flexible  Round-Ring  1" Capacity  Letter  No  No  No  Blue  5 Per Set
Flexible  Round-Ring  1" Capacity  Letter  No  No  No  Blue  5 Per Set
Flexible  Round-Ring  1" Capacity  Letter  No  No  No  Blue  5 Per Set
Flexible  Round-Ring  1" Capacity  Letter  No  No  No  Blue  5 Per Set
```

**Standardized Description**

```
Flexible: 5-Ring-Ring 1" Capacity  Black Acetate
Flexible: 5-Hole Ring 1" Capacity  Black Acetate
Flexible: 5-Hole Ring 1" Capacity  Black Acetate
Flexible: 5-Hole Ring 1" Capacity  Black Acetate
Metal: Handle + Channel: 42-7/8" of Double Hanging Data Binder, 1" Black
Metal: 5-Spring Label Holders for 15" and 1" Binders: Clear Vinyl, 12-Pack
Metal: 5-Spring Label Holders for 7-1/4" and 2" Binders: Clear Vinyl, 12-Pack
Wide: 220 Ring Presentation View Binder, 1" Capacity: Letter, White
Flexible: Double-Fastener Centers for Binding, 8-1/2" x 11" Black, 5 Per Set
```

**Silver Creek Systems ©2009**
Step 2 – Semantic Transformation to Target Format

Simple standardization
- Any attributes in any form
- Imperial/metric conversions
- Multiple classifications

Auto-translation
- Translate millions or rows in seconds
- Full double-byte support

Auto-abbreviation
- Intelligent abbreviation to meet target length
- Optimizes each line with no manual effort

Next-Generation Tools for Product Data Quality

- Semantic-based
- Auto-learning
- Integrated Governance
Next-Generation Technology

3rd Generation: Auto-Learning/Inference
- Built by: System – assisted by non-technical SME
- Method: Semantic (based on context)
- Reuse: Very good (generalizes well)
- Timeline: Hours

2nd Generation: Visual rules
- Built by: Subject Matter Expert
- Method: Semantic (based on context)
- Reuse: Very good (generalizes well)
- Timeline: Months

1st Generation: Coded rules
- Built by: IT
- Method: Syntactic (based on patterns)
- Reuse: Poor (does not generalize)
- Timeline: Years

Semantics – the identification of meaning

Underlying Semantic Model
1. Recognizes non-standard data
2. Extracts and standardizes important information
3. Learns from real data during operation
AutoBuild – Generating the Semantic Model

From Silver Creek

Generates semantic model from available metadata
Semantic model – ready to put into production

Avoids days or weeks of iterative coding & testing - per product category

Silver Creek Systems ©2009

Semantic Recognition

Recognize, Parse, Extract & Standardize

Description

<table>
<thead>
<tr>
<th>Item Type</th>
<th>Size</th>
<th>Materials</th>
<th>Grip</th>
<th>Brand</th>
<th>Length</th>
<th>Shipping Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgical</td>
<td>Large</td>
<td>Latex</td>
<td>Smooth</td>
<td>Protegrity</td>
<td>8.0 IN</td>
<td>5.1 LBS</td>
</tr>
<tr>
<td>Surgical</td>
<td>Large</td>
<td>Latex</td>
<td>Textured</td>
<td>Protegrity</td>
<td>8.0 IN</td>
<td>5.1 LBS</td>
</tr>
<tr>
<td>Surgical</td>
<td>Large</td>
<td>Latex</td>
<td>Smooth</td>
<td>Protegrity</td>
<td>9.0 IN</td>
<td>5.1 LBS</td>
</tr>
<tr>
<td>Surgical</td>
<td>Medium</td>
<td>Latex</td>
<td>Smooth</td>
<td>Protegrity</td>
<td>7.9 IN</td>
<td>5.1 LBS</td>
</tr>
<tr>
<td>Surgical</td>
<td>Medium</td>
<td>Latex</td>
<td>Textured</td>
<td>Protegrity</td>
<td>5.1 LBS</td>
<td>5.1 LBS</td>
</tr>
<tr>
<td>Surgical</td>
<td>Large</td>
<td>Latex</td>
<td>Textured</td>
<td>Protegrity</td>
<td>5.1 LBS</td>
<td>5.1 LBS</td>
</tr>
<tr>
<td>Exam</td>
<td>Small</td>
<td>Latex</td>
<td>Smooth</td>
<td>Protegrity</td>
<td>7.5 IN</td>
<td>17.5 LBS</td>
</tr>
<tr>
<td>Exam</td>
<td>Small</td>
<td>VINYL</td>
<td>Textured</td>
<td>Instagard</td>
<td>6.5 IN</td>
<td>17.5 LBS</td>
</tr>
<tr>
<td>Exam</td>
<td>Medium</td>
<td>Nitrile</td>
<td>Smooth</td>
<td>Protegrity</td>
<td>9.0 IN</td>
<td>9.0 IN</td>
</tr>
</tbody>
</table>

Highly variable syntax (patterns)
Normalized master data

Semantic Recognition handles any data

- Highly variable patterns
- Vocabulary variations – both ‘taught’ & system generated
- Punctuation and word order variations
- Ambiguities – disambiguate through semantic context
- Exceptions – identify & remediate

Silver Creek Systems ©2009
Semantic Recognition is the Key to Control...

Semantic recognition enables:
- Standardize attributes – any standard
- Standardize descriptions – any length or standard
- Translate – from any language to any language
- Classify – to any schema – UNSPSC, eClass, custom etc.
- Validate – attribute values
- Enrich – from external sources
- Match – ID match, functional match etc.
- Merge – based on survivorship rules
- Identify gaps – for enrichment
- Suggest improvements – recognition, process etc.

Tough/impossible to replicate in a traditional DQ approach

Knowledge for a single category

2,000 categories X = Way too much to build with code...
Summary

**Product Data is Different**
- Variable data
- Variable categories
- Variable uses

Traditional DQ approaches rarely work well with product data
- Wrong technology basis
- Insufficient solution scope (missing capabilities)

New technologies are emerging to meet Product Data Quality needs
- Semantic-based
- Auto-learning
- Integrated Governance