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

Silver Creek Systems ©2009

MDM Segment Size & Growth (Gartner, Nov 08)

Customer Data Quality

	2007	5 Year CAGR
Customer (CDI)	$335m	21%
Product (PIM)	$401m	22%
Procurement	$36m	29%
Asset	$56m	28%

Product Data Quality

Silver Creek Systems ©2009

silvercreek SYSTEMS

silvercreek SYSTEMS

Product Data is *Different*

- Understanding the Product Data Problem
- Comparison with 'traditional' Customer Data Quality

www.silvercreeksystems.com

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 Gartner

Silver Creek Systems ©2009

silver creek SYSTEMS

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

Item	Motor
Classification	26101600
Power	10 horsepower
Voltage	115
Mounting	Yoke

How many different ways can you receive information?

How many ways do you need it?

Silver Creek Systems ©2009

silver creek SYSTEMS

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

Product ID	Manufa-cturer	Description	Product Type	Power	Voltage	Mount
ABC123	AA Inc.	10hp motor 115V Yoke mount	Motor, AC/DC	10hp	115V	Yoke
abc-123	A.A.	mtr, ac(115) 10 horsepower 115volts	AC/DC Motor	10	115 AC.	
ABC/123/Q	AA/Craft	10 Caballos, Motor, 115 Voltios	Mot-AC	10 H-pow	115	
QA-ST5	Craft	TEAO HP = 10.0 1725RPM 115V 48YZ YOKE MTR	26101604			
Z99	Z99	MOT-10,115V,48YZ,YOKE	Z99			
		Motor powered pulley, 3/4"	Motor			

Widespread duplication (often hard to spot)

Inconsistent formats (extra characters often added)

Mis-classified item (not a motor)

Inconsistent classifications & misclassifications

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

VENTANA RESEARCH

'PIM Business & Technology Trends - Survey', Sept 2007

Silver Creek Systems ©2009

silvercreek SYSTEMS

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

Silver Creek Systems ©2009

silvercreek SYSTEMS

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'**

VENTANA RESEARCH

'PIM Business & Technology Trends - Survey', Sept 2007

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

PREMIER

AVNET®

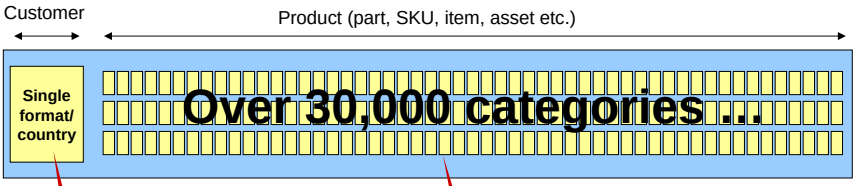
Leading hospital group

Silver Creek Systems ©2009

silver creek SYSTEMS

Customer Data Quality vs. Product Data Quality

Different problems, different technologies



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

silver creek SYSTEMS

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

Silver Creek Systems ©2009

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

AMR Research - MDM Strategies for Enterprise Applications, April 2007

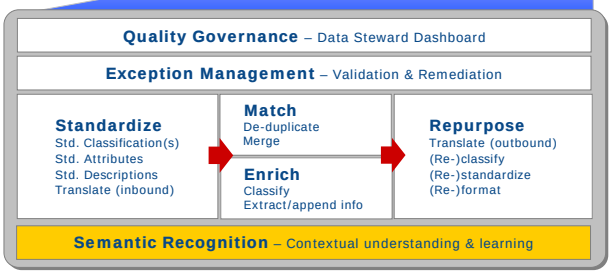
Silver Creek Systems ©2009

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



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

silvercreek SYSTEMS

Silver Creek Systems ©2009

silvercreek SYSTEMS®

Product Data Quality – Applied Use Cases

- Leading Healthcare Alliance
- World-class Retailer

www.silvercreeksystems.com

silvercreek SYSTEMS

PREMIER

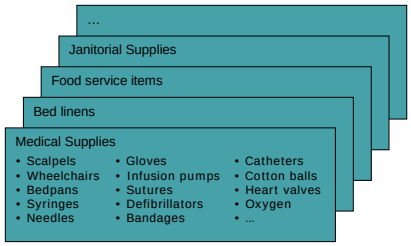
Leading Healthcare Alliance – Premier Inc.

Excerpt of Presentation at Gartner MDM Conference, November 2009

www.silvercreeksystems.com

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…



PREMIER Transforming Healthcare Together

Data Challenges – How Many Ways Can You Say "3M"?

Manufacturer	Manufacturer	Manufacturer	Manufacturer	Manufacturer
SOUTHLAND TECHNOLOGY 3M	3M 800-327-5360	3M C/O EDUMEX	3M MED. SURGICAL SALES/MICROSC	3M CARDIOVASCULAR
3M CO PHOTO PRODUCTS DIV	3M CO	3M C/O FLUID PACKAGING	3M MEDICA	ESPAME
3M DIAGNOSTIC SYSTEMS INC	3M DENTAL 800-237-1650	3M C/O GORE(X) INC	3M MEDICAL	ESPE AMERICA IN
3M ELECTRICAL SPECIALTIES DIV	3M ESPE DENTAL DIVISION 800-364-3577	3M C/O HORIZON MEDICAL	3M MEDICAL &SURGICAL	ESPE AMERICA INC
3M HEALTH	3M ESPE UNITED STATES	3M C/O INNOVATIVE TECH. LTD.	3M MEDICAL ADVANCED WC	ESPE AMERICA, INC
3M HEALTH CARE CDI	3M ESPE	3M C/O JAMES ALEXANDER CORP.	3M MEDICAL AND SURGICAL PROD.	3M PUERTO RICO
3M HEARING COMPONENTS	3M HEALTH CARE 800-521-2818	3M C/O LUDLOW TECH. PRODUCTS	3M MEDICAL SURGICAL	3M SPECIALITY CHEMICAL
3M INDUSTRIAL TAPES LTD	3M HEALTHCARE PRODUCT	3M C/O NAMSA	3M MEXICO / CLUDAD JUAREZ	3M C/O BERKLEY SURGICAL
3M MEDICAL DEVICE DIV	3M HEALTHCATE	3M C/O NORTH SAFETY PRODUCTS	3M MEXICO/CD JUAREZ	3M CANADA INC
3M MEDICAL IMAGING SYSTEMS DIV	3M MEDISURGE	3M C/O OVERLAND BOLLING	3M MINNESOTA MINI	3M C/O MITSUI TEXTILE CORP
3M MEDICAL PRODUCTS DIV	3M MINNESOTA MINING MFG OFFICE	3M C/O REMINGTON PRODUCTS, INC	3M MINNESOTA MINING & MFG. CO.	3M PHARMACEUTICAL SUB OF MINNESOTA
3M MEDICAL-SURGICAL DIV	3M MINNESOTA MINNING & MFG CO.	3M C/O SAN-MAR LAB, INC.	3M MINNESOTA MINING & MFG CO	THREE M
3M MEDICALSURG	3M OCC. HEALTH AND ENV. SAFETY DIV	3M C/O SCAPA TAPES/RENFREW	3M MINNESOTA MINING & MFG CO.	THREE M (3M)INFUSION THERAPY
3M PHARMACEUTICS AND MEDICAL S	3M OCC. HEALTH AND ENV. SAFETY DIV.	3M C/O THAI HOSP PROD CO. LTD.	3M MINNESOTA MINING AND MFG	THREE M COMPANY
3M-MEDICAL/SURGICAL	3M SARNI/CDI	3M C/O THAI HOSPITAL PRODS CO.	3M OCCUPATIONAL HEALTH	THREEM
3M/ OCCUPATIONAL AND SAFETY DIV	3M SURGICAL	3M C/O THAI HOSPITAL PRODUCTS	3M OCCUPATIONAL HEALTH AND ENV	THREEMESPE
3M - MINNESOTA MINING &	3M UNITEK 800-423-4588	3M C/O TIVERSICO	3M OEM KIT PRODUCTS	5 P RICHARDS
CO	3M UNITEK	3M C/O W. PHARMACEUTICAL SVCS	3M OFFICE SUPPLY & HOUSEKEEPING CO	MINNESOTA MINING & M
3M FEDERAL GOVERNMENT	THREE MI ESPE	3M C/O WAHL CORP	3M OH & ES DIVISION PLANT	MINNESOTA MINING & MANUFACTURI
3M FEDERAL SYSTEMS DEPARTMENT	3M COMPANY C/O WAHL CORP.	3M COMMERCIAL OFFICE SUPPL	3M ORTHOPEDIC	MINNESOTA MINING & MFG
3M HEALTH CARE SYSTEMS	MINNESOTA SCIENTIF	3M COMMERCIAL OFFICE SUPPLY	3M PHARM	MINNESOTA MINING & MFG
3M HEALTHCARE $250 MINIMUM ORDER	CORPORATE ALLIANCE 3M CUSTOMER SERV	3M COMPANY	3M PHARMACEUTICAL	MINNESOTA MINING AND MANUFACT
3M HEALTHCARE(MINNSOTA MINNING)	3 M HEALTHCARE	3M COMPANY ABERDEEN PLANT	3M PHARMACEUTICAL SUB OF MINNESOTA	MINNESOTA MINING AND MFG CO
3M MEDICAL - CREDIT CARD	3 M UNITEK CORP	3M COMPANY BROOKINGS PLANT	3M PHARMACEUTICAL SUB OF MINNESOTA M	MINNESOTA MINING AND MFG CO 3M DENTA
3M MEDICAL PRODUCTS	3-M	3M COMPANY C/O EDUMEX	3M PHARMACEUTICALS	MINNESOTA MINING AND MFG CO ABRASIVE
3M OCC. HEALTH AND ENV. SAFETY DIVISION	3-M COMPANY	3M COMPANY IRVINE PLANT	3M PHARMACEUTICALS INC	MINNESOTA MINING AND MFG CO ADHESIVE
3M OCCUPATIONAL AND SAFETY DIV	3-M COMPANY-C/O O	3M COMPANY, BROOKINGS SD	3M PODIATRY	MINNESOTA MINING AND MFG CO DENTAL P
3M SAFETY DIVISION	3-M COMPANY-C/O OEM PRODUCTS	3M CONSTRUCTION MARKETS	3M PROF HEALTH	MINNESOTA MINING AND MFG CO DIAGNOST
3M-DENTAL PRODUCTS DIVISION	3-M PHARMACEUTICALS	3M CONSUMER HEALTH CARE, C/O 3	3M SURGICAL DIV STAPLIN	MINNESOTA MINING AND MFG CO ELECTRIC
3M-HEALTH CARE	33M HEALTHCARE	3M CORPORATE ALLIANCE	3M TEGADERM	MINNESOTA MINING AND MFG CO HEALTH C
3M DENTAL PRODUCTS DIV	3M	3M DENTAL	3M UNITED KINGDOM PLC	MINNESOTA MINING AND MFG CO INDUSTRI
3M UNITEK CORPORATION	3M PUERTO RICO	3M DENTAL PRODUCTS	3M UNITEK CORP	MINNESOTA MINING AND MFG CO INTL EXP
3M UNITEK DENTAL PRODUCTS	3M SPECIALITY CHEMICAL	3M DEUTSCHLAND GMBH	3M-MINNESOTA MINI	MINNESOTA MINING AND MFG CO MEDICAL
3M BIOLOGICAL	3M % SAN-MAR	3M HEALTCARE	3M-MINNESOTA MINING & MFG CO	MINNESOTA MINING AND MFG CO OCCUPAT
3M ESPE DENTAL PRODUCTS	3M (CRUTZ42)	3M HEALTH CARE	3M-MINNESOTA MINING & MFG. CO	MINNESOTA MINING AND MFG CO PACKAGIN
3M HEALTH CARE (MEDISURG PRODS)	3M - MINNESOTA MI	3M HEALTH CARE LTD	3M-MINNESOTA MINING & MFG. CO.	MINNESOTA MINING AND MFG CO PERSONA
3M C/O CHECKPOINT METO	3M - MINNESOTA MINING & MFG.CO.	3M HEALTH CARE LTD.	3M-MINNESOTA MINING & MFG.CO.	MINNESOTA MINING AND MFG CO SALES CE
3M CANADA PLANT	3M -MINNESOTA MINING & MFG. CO.	3M HEALTH CARE SERVICE CENTER	3M-MINNESOTA MINNING & MFG. CO.	MINNESOTA MINING AND MFG CO STATIONE
3M HEALTH CARE PLANT	3M -OHMEDS DIVISION PLANT	3M HEALTH CARE(MEDICAL/SURGICA	3M-NAT ACCTS SALES	MINNESOTA MINING AND MFG CO SURGICAL
MINNESOTA MINING AND MFG	3M AND COLOPLAST	3M HEALTH INFORMATION SYSTEMS	3M-NAT ACCTS SALES DIST	MINNESOTA MINING COMMERICAL CARE
MINNESOTA MINING AND MFG.	3M BUSINESS PRODUCTS	3M HEALTHCARE	3M-NAT ACCTS SALES(OBSOLETE)10-99	MINNESOTA MINING, 3M
MINNESOTA MINNING AND MFG.	3M C/O ACCUPAC	3M HEALTHCARE OEM	3MMED	MINNESOTA SCIENTIFIC INC
CORPORATE ALLIANCE	3M C/O ACCUPAC INC	3M HEALTHCARE PRODUCTS	3 M UNITEK CORPORATION	MINNESOTA-MINING AND MANUFACT
3M HEALTHCARE	3M C/O ALBA WALDENSIAN	3M HOME & COMMERICAL CARE DIV	CORPORATE ALLIANCE 3M CUSTOMER SERV	3M FRANCE
3M - SCOTCH	3M C/O CELESTE INDUSTRIES	3M INFUSION THERAPY/WI	CORPORATE ALLIANCE 3M CUSTOMER SERVI	ALLERGENETICS DIV AXIONICS INC
				CARDIOVASCULAR DEVICES
				3M HEALTHCARE/MAYO ONLY

PREMIER Transforming Healthcare Together

Data Challenges – Same Product, Different Numbers

Different distributors and hospitals use different identification numbers

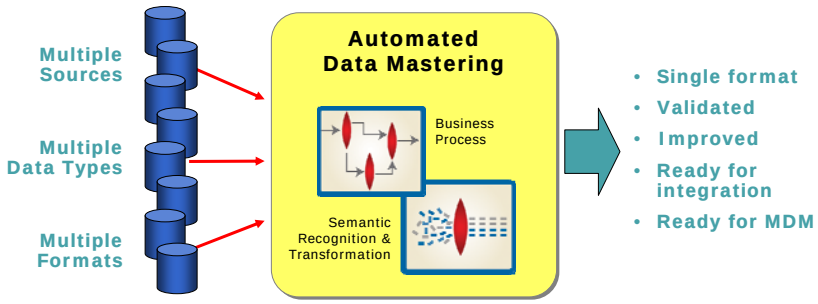
Allegiance	**- M8630**
Owens & Minor	**- 4509008630**
BBMC-Colonial	**- 045098630**
BBMC-Durr	**- 081048**
Kreisers	**- MINN8630**
Midwest	**- TM-8630**
Pacific	**- 3/M8630**
UnitedUMS	**- 001880**

Industry Distributor Numbers for 3M Product # 8630:

3M™ DuraPrep™ Surgical Solution (Iodine Povacrylex [0.7% available Iodine] and Isopropyl Alcohol, 74% w/w) Patient Preoperative Skin Preparation

PREMIER Transforming Healthcare Together

What is Data Mastering?



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

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

PREMIER Transforming Healthcare Together

ID and Semantic Matching

Standardized to Matched to

	From the PO	Standardized to	PIM Master Data
Description	REAGENT METHOTREXATE II		REAGENT TDM METHOTREXATE II FPIA
Manufacturer	ABBOTT DIAGNOSTIC	ABBOTT LABORATORIES, INC.	ABBOTT TDX/TDXFLX 100TEST
Part Number	07A12-60	07A1260	ABBOTT LABORATORIES, INC.
UOM	EA	EA	07A1260
			EA

Standardize manufacturer; drop PN special characters

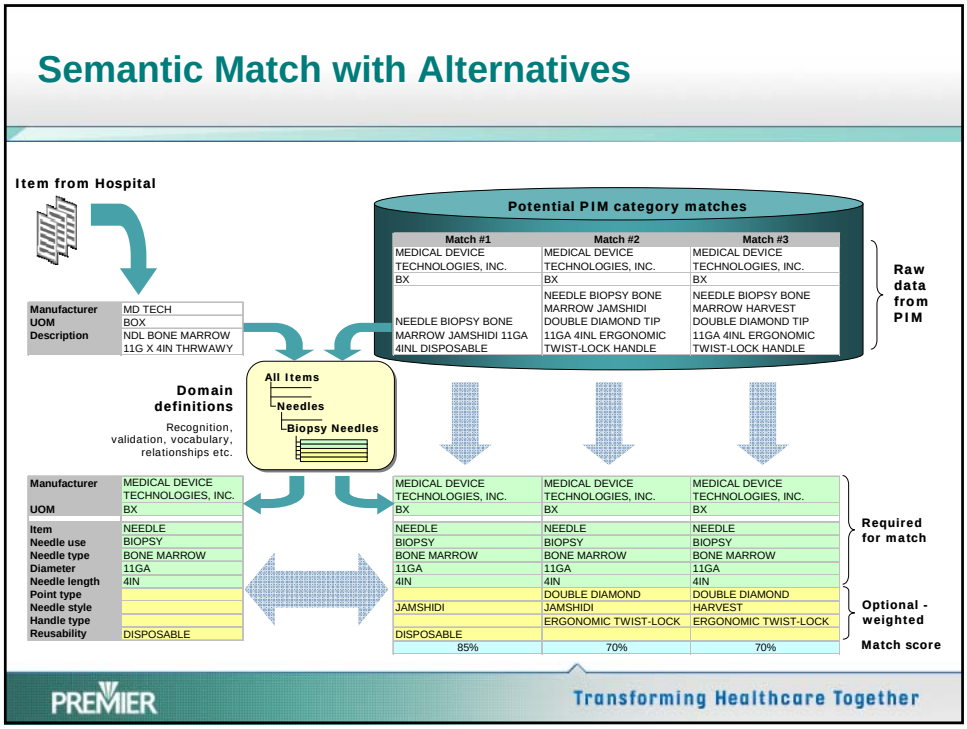
	From the PO	Standardized to	PIM Master Data
Description	NEEDLE 15-DEGREE STAMEY 095015 ANNUAL		NEEDLE UROLOGY STAMEY 15DEG 6 3/4IN
Manufacturer	COOK INC	COOK GROUP, INC.	COOK UROLOGICAL INCORPORATED
Part Number	G15081	095015	**Mfg Family:** COOK GROUP, INC.
UOM	EA	EA	095015
			EA

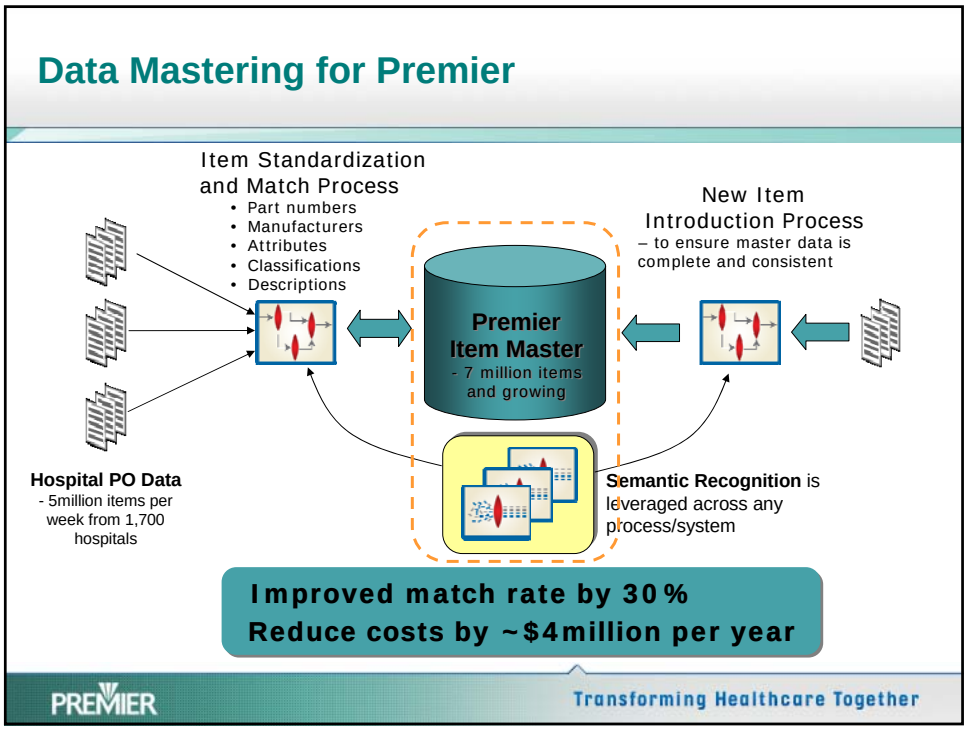
Standardize manufacturer; find Mfg parent match; find PN in descriptive text

	From the PO	Standardized to	PIM Master Data
Description	NDL ANASTOMIC DVC 8PK S20UCLIP		DEVICE ANASTOMOSIS TITANIUM CLIP SINGLE ARM SINGLE NEEDLE 20FR ID SHORT FLEX REGULAR NEEDLE 8/PK
Manufacturer	MEDTRONIC USA INC	MEDTRONIC USA, INC.	MEDTRONIC USA, INC.
Part Number	S20SFRN8	S20SFRN8	S20SFRN8
UOM	BX	PK	PK

Standardize manufacturer; correct UOM from descriptive text

PREMIER Transforming Healthcare Together


Semantic Match with Alternatives
Item from Hospital
Manufacturer MD TECH
UOM BOX
Description NDL BONE MARROW 11G X 4IN THRWAWY
Potential PIM category matches
Match #1 MEDICAL DEVICE TECHNOLOGIES, INC. | BX | NEEDLE BIOPSY BONE MARROW JAMSHIDI 11GA 4INL DISPOSABLE
Match #2 MEDICAL DEVICE TECHNOLOGIES, INC. | BX | NEEDLE BIOPSY BONE MARROW JAMSHIDI DOUBLE DIAMOND TIP 11GA 4INL ERGONOMIC TWIST-LOCK HANDLE
Match #3 MEDICAL DEVICE TECHNOLOGIES, INC. | BX | NEEDLE BIOPSY BONE MARROW HARVEST DOUBLE DIAMOND TIP 11GA 4INL ERGONOMIC TWIST-LOCK HANDLE
Raw data from PIM
Domain definitions
Recognition, validation, vocabulary, relationships etc.
All Items
Needles
Biopsy Needles
Manufacturer MEDICAL DEVICE TECHNOLOGIES, INC.
UOM BX
Item NEEDLE
Needle use BIOPSY
Needle type BONE MARROW
Diameter 11GA
Needle length 4IN
Point type
Needle style
Handle type
Reusability DISPOSABLE
MEDICAL DEVICE TECHNOLOGIES, INC. | BX | NEEDLE | BIOPSY | BONE MARROW | 11GA | 4IN | JAMSHIDI | DISPOSABLE
MEDICAL DEVICE TECHNOLOGIES, INC. | BX | NEEDLE | BIOPSY | BONE MARROW | 11GA | 4IN | DOUBLE DIAMOND | JAMSHIDI | ERGONOMIC TWIST-LOCK
MEDICAL DEVICE TECHNOLOGIES, INC. | BX | NEEDLE | BIOPSY | BONE MARROW | 11GA | 4IN | DOUBLE DIAMOND | HARVEST | ERGONOMIC TWIST-LOCK
Required for match
Optional - weighted
85% 70% 70%
Match score
PREMIER
Transforming Healthcare Together


Data Mastering for Premier
Item Standardization and Match Process
• Part numbers
• Manufacturers
• Attributes
• Classifications
• Descriptions
New Item Introduction Process
– to ensure master data is complete and consistent
Premier Item Master
- 7 million items and growing
Hospital PO Data
- 5million items per week from 1,700 hospitals
Semantic Recognition is leveraged across any process/system
Improved match rate by 30%
Reduce costs by ~$4million per year
PREMIER
Transforming Healthcare Together

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

PREMIER Transforming Healthcare Together

silvercreek SYSTEMS

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

Count of Level 3	Dimensions										
#SKUS - range	0	1	2	3	4	5	6	7	8	9	Grand Total
<25	297	36	22	39	16	3	2		1		416
25-49	152	29	15	28	7	2	6	3	4		246
50-74	41	6	9	8	2	3	2	1	3		75
75-99	25	14	2	1	2			3		1	48
>=100	12	4		4	2	1	2	1			26
Grand Total	527	89	48	80	29	9	12	8	8	1	811
Blindspots	230	24	2	4							260

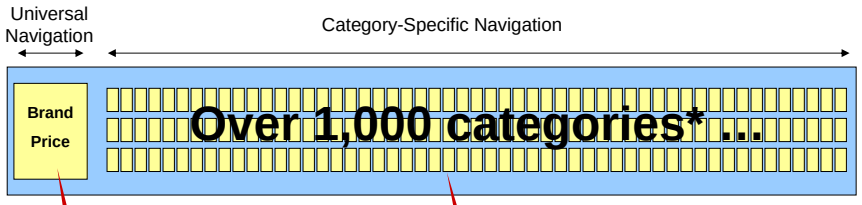
- 65% of categories have no category-specific navigation
- Half of these categories have > 25 SKUs

Poor website usability

Silver Creek Systems ©2009

silvercreek SYSTEMS

The Challenge: Category-specific Information



Universal Dimensions

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

BUT Rarely adequate for "World class" guided navigation

Category-specific Dimensions

- 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

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

Silver Creek Systems ©2009

*typical retail scenario

silvercreek SYSTEMS

Category-Specific Information

Silver Creek Systems ©2009

Attributes must be extracted and standardized

Short Description	Disney - Tinker Bell Metallic Tote Bag
Bullets	Novelty graphics Zippered main compartment Zippered interior pocket Double O-ring straps Measures approximately: 12"W x 11"H x 4"D; 9" drop Imported fuschsia

Attribute	Extracted	Standardized
Item_Name	Tote Bag	Tote Bag
Brand or Model Name	Disney	Disney
Characters	Tinker Bell	Tinker Bell
Size	12"W x 11"H x 4"D	Large
Color	Fuschia	Pink
Closure	Zippered main compartment	Zippered
Strap type	Double O-ring straps	Dual Straps
Drop length	9"	9 to 11 inches

Category-specific extraction and standardization

Sample Standardizations

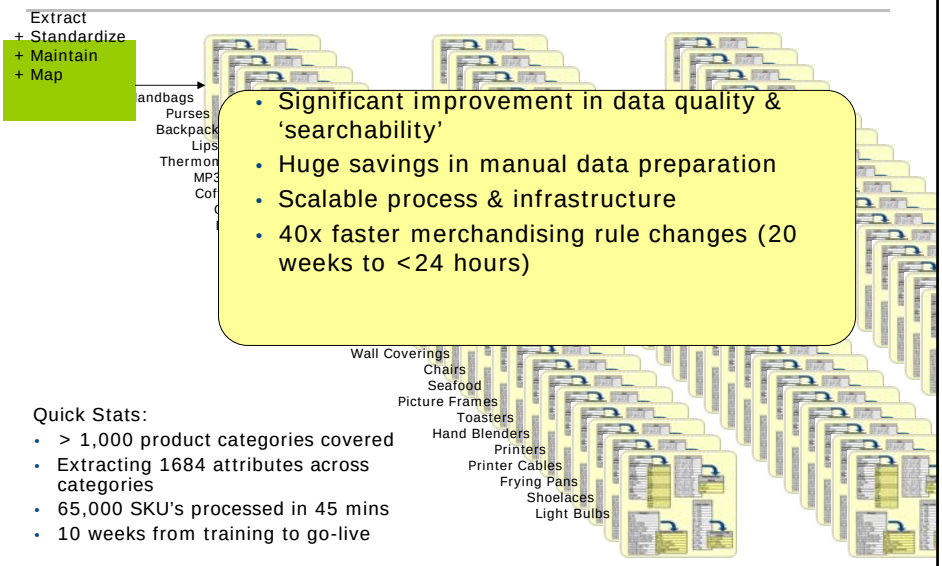
Colors	Standardized Colors
apple	Green
foamy aqua	Blue
asparagus	Green
Bronze	Brown
cement	Gray
Cognac	Brown
eggplant	Purple
emerald	Green
Expresso	Brown
faded yellow	Purple
fuchsia	Yellow
Ivory	Pink
jade	White
khaki	Green
lavander	Brown
bright lemon	Purple
	Yellow

Character	Standardized Characters
Cuties	Cuties
Eeyore	Eeyore
Grumpy	Grumpy
hannah montana	Hannah Montana
Hannah Montana	High School Musical
High School Musical	Mickey Mouse
HighSchool Musical	SpongeBob SquarePants
Mickey Mouse	Tinker Bell
Mickey Mouse East/West	
Sponge Bob Square Pants	
SpongeBob SquarePants	
Tinker Bell	
Tinker Bell East/West	
Tinker Bell Pop	
Tinker Bell Shimmer	
Tinker Bell Sketch	

Drop Length	Standardized Drop Length
10" drop	Under 6 inches
11" drop	6 to 8 inches
12" drop	9 to 11 inches
13" drop	12 inches or More
19" drop	
20" drop	
4.5" drop	
5" drop	
5.5" drop	
6" drop	
7" drop	
7.5" drop	
8" drop	
8.5" drop	
9" drop	
9.5" drop	

Silver Creek Systems ©2009

Live with 1,000 product categories in 10 weeks



- 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

Silver Creek Systems ©2009

silvercreek SYSTEMS

Semantic Recognition & Transformation Examples

Step 1 – Semantic Recognition of Category

Input Description
4" end **cap**
4 in **cap** screw
4 in box b-ball **cap**
4 ESU **cap**
750ml Merlot, screw **cap**

DataLens

Category
Piping & plumbing
Fasteners
Headgear
Capacitors
Wine

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

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

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

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

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

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

Silver Creek Systems ©2009

silver creek SYSTEMS

Step 2 – Semantic Transformation to Target Format

Sample: Resistors

ITEM DESCRIPTION
RESPCF AX3000HM1/4W5%
220 ohm CF AX 0.5W 5 % ress
array 16 pin 85 ohm 5% resp
NTWK 330 OMH 8 P 2% RESP
CF, AX 75 OHM 1/4 W 5% RESP
NTWK 25P 220 OHM RESS
RES MF AX 226 OHM 1/4 W 1%

Highly variable syntax (patterns)

DataLens

Resist-ance	Power	Toler-ance
300 Ohm	0.25 Watt	5%
220 Ohm	0.5 Watt	5%
85 Ohm		5%
330 Ohm		2%
75 Ohm	0.25 Watt	5%
220 Ohm		
226 Ohm	0.25 Watt	1%

Key attributes extracted and normalized

UNSPSC	FSC
32121609	5905
32121609	5905
32121607	5905
32121607	5905
32121609	5905
32121607	5905
32121609	5905

Classified to many number of schemas – industry standard or custom

Standard Description
Resistor 300 Ohm 5% 0.25 Watt Axial Carbon Film
Resistor 220 Ohm 5% 0.5 Watt Axial Carbon Film
Resistor 85 Ohm 5% 16 Pin Array
Resistor 330 Ohm 2% 8 Pin Network
Resistor 75 Ohm 5% 0.25 Watt Axial Carbon Film
Resistor 220 Ohm 25 Pin Network
Resistor 226 Ohm 1% 0.25 Watt Axial Metal Film

Normalized descriptions to fit any standard or format

Russian Description
Резистор 300 Ом 5% 0.25 Ватт Аксиальная Карбонная Пленка
Резистор 220 Ом 5% 0.5 Ватт Аксиальная Карбонная Пленка
Резистор 85 Ом 5% 16 Выводов Массив
Резистор 330 Ом 2% 8 Выводов Сеть
Резистор 75 Ом 5% 0.25 Ватт Аксиальная Карбонная Пленка
Резистор 220 Ом 25 Выводов Сеть
Резистор 226 Ом 1% 0.25 Ватт Аксиальная Металическая Пленка

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

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

Item can then be cleansed (standardized) as desired

Silver Creek Systems ©2009

silver creek SYSTEMS

Step 2 – Semantic Transformation to Target Format

ORIGINAL DESCRIPTION
TEAO, 1/4HP,1725RPM,115V,48YZ,YOKE, MTR
,1725RPM MTR, 1/4HP115V,48YZ,YOKE, TEAO
TEAO Motor, YOKE, 1/4HP,1725RPM,115V,48YZ
MOTOR,TEAO, 1/2HP,1075RPM,115V,48YZ,YOKE
1075RPM,115V,MOTOR,TEAO, 1/2HP,48YZ,YOKE

Sample: Motors

Highly variable syntax (patterns)

DataLens™

Encl-osure	Horse Power	Speed	Frame Size	Voltage	Mounting
TEAO	1/4 HP	1725 RPM	48YZ	115 V	Yoke
TEAO	1/4 HP	1725 RPM	48YZ	115 V	Yoke
TEAO	1/4 HP	1725 RPM	48YZ	115 V	Yoke
TEAO	1/2 HP	1075 RPM	48YZ	115 V	Yoke
TEAO	1/2 HP	1075 RPM	48YZ	115 V	Yoke

Key attributes extracted and normalized

UNSPSC Category	UNSPS C Desc
26101600	Motors
26101600	Motors
26101600	Motors
26101600	Motors
26101600	Motors

Classified to many number of schemas – industry standard or custom

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

Item can then be cleansed (standardized) as desired

Standard Description
Motor,TEAO,1/4 Horse Power,1725 RPM,48YZ,115 Volts,Yoke
Motor,TEAO,1/4 Horse Power,1725 RPM,48YZ,115 Volts,Yoke
Motor,TEAO,1/4 Horse Power,1725 RPM,48YZ,115 Volts,Yoke
Motor,TEAO,1/2 Horse Power,1075 RPM,48YZ,115 Volts,Yoke
Motor,TEAO,1/2 Horse Power,1075 RPM,48YZ,115 Volts,Yoke

Normalized descriptions to fit any standard or format

Spanish Description
Motor,TEAO,Potencia de 1/4 Caballos,1725 RPM,48YZ,115 Voltios,Montaje de Yugo
Motor,TEAO,Potencia de 1/4 Caballos,1725 RPM,48YZ,115 Voltios,Montaje de Yugo

Russian Description
мотор,TEAO,1/4 Лошадиных Сил,1725 Оборотов в минуту,48YZ,115 Вольт,хомут
мотор,TEAO,1/4 Лошадиных Сил,1725 Оборотов в минуту,48YZ,115 Вольт,хомут
мотор,TEAO,1/4 Лошадиных Сил,1725 Оборотов в минуту,48YZ,115 Вольт,хомут
мотор,TEAO,1/2 Лошадиных Сил,1075 Оборотов в минуту,48YZ,115 Вольт,хомут
мотор,TEAO,1/2 Лошадиных Сил,1075 Оборотов в минуту,48YZ,115 Вольт,хомут

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

Silver Creek Systems ©2009

silvercreek SYSTEMS

Step 2 – Semantic Transformation to Target Format

PRODUCT DESCRIPTION
Flexible Accohide 3-Round-Ring Binder, 1" Capacity, 8-1/2"x11", Black
Flexible Accohide 3-Round-Ring Binder, 1" Capacity, 8-1/2"x11", Dark Royal Blue
Flexible Accohide 3-Round-Ring Binder, 1" Capacity, 8-1/2"x11", Blue
Accodata Handle + Channel, 14-7/8" or Smaller Hanging Data Binder, 17"L, Black
Ident-O-Spine Label Holders for 1/2" and 1" Binders, Clear Vinyl, 12/Pack
Ident-O-Spine Label Holders for 1-1/2" and 2" Binders, Clear Vinyl, 10/Pack
X-Wide EZD-Ring Presentation View Binder, 1" Capacity, Letter, White
Double Pocket Dividers for 3-Ring Binders, 8-1/2"x11", Buff, 5/Set

Sample: Binders

DataLens™

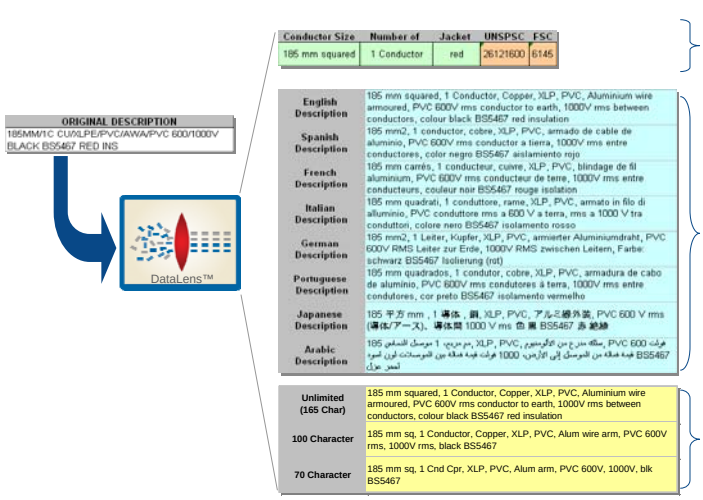
Type	Ring Type	# Rings	Sheet Capacity	Sheet Size	Cover Material	Lock-ing	Inside pocket	Label Holder	Color	Packaging
Flexible	Round-Ring		1 inch Capacity	Letter		No	No	No	Black	
Flexible	Round-Ring		1 inch Capacity	Letter		No	No	No	Dark Royal Blue	
Flexible	Round-Ring		1 inch Capacity	Letter		No	No	No	Blue	
Hanging						No	No	No	Black	
-----					Vinyl	No	No	No	Clear	12 Per Pack
-----					Vinyl	No	No	No	Clear	10 Per Pack
View	D-Ring		1 inch Capacity	Letter		No	No	No	White	
-----		3-Ring		Letter		No	No	No		5 Per Set

STANDARDIZED DESCRIPTION
Binder, Round-Ring Flexible 3 1 inch Capacity 8-1/2"x11" black Accohide -,,,
Binder, Round-Ring Flexible 3 1 inch Capacity 8-1/2"x11" dark royal blue Accohide -,,,
Binder, Round-Ring Flexible 3 1 inch Capacity 8-1/2"x11" blue Accohide -,,,
Binder, Hanging Data 14-7/8" 17 black Accodata Handle + Channel , or Smaller, "L ,
Binder, Vinyl 1/2" 1" clear 12 Per Pack Ident-O-Spine Label Holders for and,,
Binder, Vinyl 1-1/2" 2" clear 10 Per Pack Ident-O-Spine Label Holders for and,,
Binder, EZD-Ring Presentation View 1 inch Capacity Letter white X-Wide,,,
Binder, 3-Ring 8-1/2"x11" 5 Per Set Double Pocket Dividers for, , Buff ,

Silver Creek Systems ©2009

silvercreek SYSTEMS

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

Silver Creek Systems ©2009

silver creek SYSTEMS

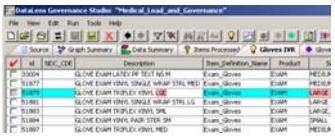
Next-Generation Tools for Product Data Quality

- Semantic-based
- Auto-learning
- Integrated Governance

www.silvercreeksystems.com

Next-Generation Technology

3rd Generation:
Auto-Learning / Inference

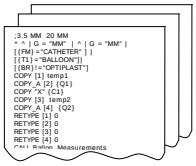


Built by: System – assisted by non-technical SME
Method: **Semantic** (based on context)
Reuse: Very good (generalizes well)
Timeline: Hours

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

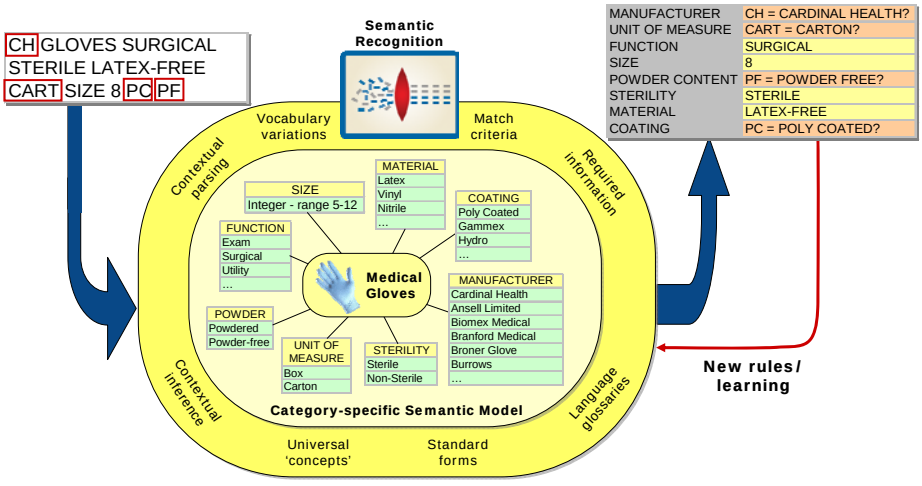
Silver Creek Systems ©2009

silver creek SYSTEMS

Semantics – the identification of *meaning*

CH GLOVES SURGICAL
STERILE LATEX-FREE
CART SIZE 8 PC PF

MANUFACTURER	CH = CARDINAL HEALTH?
UNIT OF MEASURE	CART = CARTON?
FUNCTION	SURGICAL
SIZE	8
POWDER CONTENT	PF = POWDER FREE?
STERILITY	STERILE
MATERIAL	LATEX-FREE
COATING	PC = POLY COATED?



Underlying Semantic Model

1. Recognizes non-standard data
2. Extracts and standardizes important information
3. Learns from real data during operation

Silver Creek Systems ©2009

silver creek SYSTEMS

AutoBuild – Generating the Semantic Model

From Silver Creek

Accelerator Knowledgebase

Smart Glossaries
- Dimensions & Sizes
- Materials & Colors
- Vendors & Brand names
- For sale packaging
- UNSPSC Classification scheme
- Etc.

Industry Knowledge
- Pharma standards, drug names
- NCD Codes
- Existing domain schemas
- Etc.

From Customer

ATTRIBUTE	SAMPLE VALUES
FUNCTION	EXAM, SURGICAL, UTILITY
STERILITY	
MATERIAL	LATEX, VINYL, NITRILE
POWDER CONTENT	
MATURITY	LARGE, MEDIUM, SMALL
SIZE	7, 7.5, 8
OTHER FEATURES	RADIATION RESISTANT, EXTRA PROTECTION
COLOR	BROWN, GREEN
BRAND	NEUTRALON, ALOETOUCH
COATING	POLY COATED
LENGTH	ELBOW, SHOULDER
GENDER	MENS, WOMENS
TEXTURED	
PACKAGING	SINGLE, PAIR
DIRECTION	LEFT, RIGHT

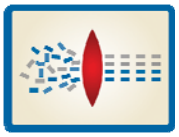
L1 CAT	L2 CAT	L3 CAT
GLOVES	CHEMOTHERAPY	LATEX POWDER-FREE
GLOVES	CHEMOTHERAPY	LATEX-FREE POWDER-FREE
GLOVES	CHEMOTHERAPY	NITRILE POWDER-FREE
GLOVES	EXAM	LATEX POWDER-FREE
GLOVES	EXAM	LATEX POWDERED
GLOVES	EXAM	LATEX-FREE

CLEANSED DESCRIPTION - SAMPLES
GLOVES SURGICAL STERILE LATEX-FREE POLY COATED POWDER-FREE SIZE 8.0
GLOVES EXAM NITRILE POWDER-FREE LARGE
GLOVES SURGICAL STERILE LATEX-FREE POLY COATED POWDER-FREE SIZE 6.0
GLOVES SURGICAL TEXTURED STERILE LATEX POLY COATED POWDER BROWN S
GLOVES SURGICAL STERILE LATEX-FREE POLY COATED POWDER-FREE SIZE 7.5

Generates semantic model from available metadata

AutoBuild

Semantic model – ready to put into production



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

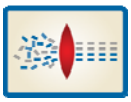
Silver Creek Systems ©2009

silver creek SYSTEMS

Semantic Recognition

Description
GLOVE PROTEGRITY SMTH GRIP MICRO PF LATEX SURG 6.0" LRG 100 / BX 5LBS
GLOVE LARGE 100EA PROTEGRITY PF LATEX SURGICAL 6.0" TEXTURE GRIP
GLVS LINER PWD 6.0" FR LTX SUG TEXT GRP ULTRATUFF CUT RESIST STRL LG
PROTE GLVS SURG LARGE PF SMTH GRIP CASE OF 1000 LTX BLU 9.0IN.
GLVE PROTGRITY TEXT SMT PF LATX SURG 7.5IN MD 250 CTN 10LBS
GLVE PROTEGRITY TEXT 100 BOX PF LATEX SURGICAL 5.5IN 20-LBS. MEDIUM
LF GLOVES TEXT ESTM SMT PF SYNTHETIC XLG SURG 7.5IN. 500 / CASE 20 LBS.
GLOV EXAM LTX 7.5IN FREE SYN (VINYL) VHA+ PF N/S SM 17.5 LBS.
XLRG GLOVES SMTH STERILE PWD LTX PROCEDURE 8.0IN
GLOVE EXAM TEXT 6.5" LTX FR N/S PWD VINYL INSTAGARD SM TEXT GRIP
GLOVE 8IN. SMTH MD NOVA PLUS PF NITRILE EXAM TEXTURE

Recognize, Parse, Extract & Standardize



Highly variable syntax (patterns)

Normalized master data

Item Type	Size	Materials	Grip	Brand	Length	Shipping Weight
Surgical	Large	Latex	Smooth	Protegrity	6.0 IN.	5 LBS.
Surgical	Large	Latex	Textured	Protegrity	6.0 IN.	
Surgical	Large	Latex	Textured	Ultratuff	6.0 IN.	
Surgical	Large	Latex	Smooth	Protegrity	9.0 IN.	
Surgical	Medium	Latex	Smooth	Protegrity	7.5 IN.	10 LBS.
Surgical	Medium	Latex	Textured	Protegrity	5.5 IN.	20 LBS.
Surgical	Extra Large	Synthetic	Textured	Esteem	7.5 IN.	20 LBS.
Exam	Small	Latex			7.5 IN.	17.5 LBS.
Procedure	Extra Large	Latex	Smooth		8.0 IN.	
Exam	Small	VINYL	Textured	Instagard	6.5 IN.	
Exam	Medium	Nitrile	Smooth	Nova Plus	8 IN.	

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

silver creek SYSTEMS

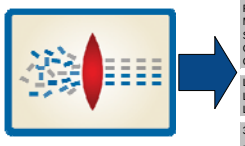
Semantic Recognition is the Key to Control…

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

GLOVES SURGICAL STERILE LATEX-FREE POLY COATED POWDER-FREE SIZE 8.0



FUNCTION	SURGICAL
STERILITY	STERILE
MATERIAL	LATEX-FREE
POWDER CONTENT	POWDER-FREE
MATURITY	LARGE
SIZE	8
COLOR	
COATING	POLY COATED

LEVEL 1 CLASS	GLOVES
LEVEL 2 CLASS	--SURGICAL
LEVEL 3 CLASS	----LATEX-FREE POWDER-FREE

30 CHARACTER	GLOVE SURG LTXF PF 8.0
50 CHARACTER	GLOVE SURGICAL STERILE LTXF POLY COAT PF SZ 8.0
FULL	GLOVES SURGICAL STERILE LATEX-FREE POLY COATED POWDER-FREE SIZE 8.0

Silver Creek Systems ©2009

silver creek SYSTEMS

Scales Across Thousands of Product Categories

Knowledge for a single category

Gloves
Catheters
Pharma - OTC
Pharma - Rx
Thermometers
Ultrasonic Gels
Forceps

2,000 categories **X**

- Extract
- Standardize
- Match/merge
- Validate
- Enrich
- Metrics

= **Way too much to build with code…**

Mops/Brushes
Needles
Liquid Soap
Leg bags
Laryngoscopes
Insufflator
Glucose Controls

Silver Creek Systems ©2009

silver creek SYSTEMS

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

Silver Creek Systems ©2009

silver creek SYSTEMS