Ball Bearings Data Analysis

# SharedLens Attribute Development

# 1. Executive Summary

**Data Overview:**

* • Total Records: 40 ball bearing items
* • Unique Part Numbers: 31 (some have multiple supplier qualifiers)
* • Total Attributes: 59 columns
* • Category: Ball Bearings (100% consistent)

**Key Findings:**

* 1. Multiple data quality issues: inconsistent UOM handling, mixed data types
* 2. Qualifier field contains supplier-specific versions (Part.SKF, Part.NTN, etc.)
* 3. Critical dimensional attributes have good fill rates (80-100%)
* 4. Engineering attributes mixed with procurement/supplier data
* 5. Unused Min/Max range columns (0% fill rate)

# 2. Data Quality Issues

## 2.1 UOM Inconsistencies

**Problem:** Load ratings stored with UOM embedded in value

* • Dynamic Load Rating: "6069.87LB" (text field, not decimal)
* • Dynamic Load Rating (KN): 27.00 (numeric field, separate)
* • Static Load Rating: "3439.6LB" (text field, not decimal)
* • Static Load (kN): 15.30 (numeric field, separate)

**SharedLens Approach:**

* Store: dynamic\_load\_rating = 27.00 (decimal, storage\_uom: "kN")
* Engineering Alias: "Dynamic Load Rating: 27.00 kN"
* Procurement Alias: "Dynamic Load Rating: 6,069.87 lbf" (via SI conversion)

## 2.2 Dimensional Data Issues

**Problem:** Inconsistent decimal precision and mixed text/numeric

* • Inside Diameter: stored as text ("2.834", "1.7717") with separate UOM column
* • Outside Diameter: stored as text ("1.378", "3.937") with separate UOM column
* • Overall Width: stored as text ("0.67", "0.98") with separate UOM column

**SharedLens Approach:**

* inside\_diameter: decimal (storage\_uom: "mm" or "in" based on source)
* outside\_diameter: decimal (storage\_uom: "mm" or "in" based on source)
* overall\_width: decimal (storage\_uom: "mm" or "in" based on source)
* All conversions handled via SI UOM system

## 2.3 Unused Attributes

**Problem:** Multiple min/max range columns with 0% fill rate

* • Max Inside Diameter + UOM (0% fill)
* • Min Inside Diameter + UOM (0% fill)
* • Max Outside Diameter + UOM (0% fill)
* • Min Outside Diameter + UOM (0% fill)

**Recommendation:** Remove unused attributes from schema or mark as optional with clear purpose documentation.

# 3. Attribute Classification for SharedLens

## 3.1 Core Engineering Attributes (Dimensional)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Current Name | SharedLens Name | Data Type | Storage UOM | Fill Rate | Key Attr? |
| Inside Diameter | inside\_diameter | Decimal | mm | 100% | Yes |
| Outside Diameter | outside\_diameter | Decimal | mm | 100% | Yes |
| Overall Width | overall\_width | Decimal | mm | 100% | Yes |

## 3.2 Performance Attributes (Load Ratings)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Current Name | SharedLens Name | Data Type | Storage UOM | Fill Rate | Key Attr? |
| Dynamic Load Rating (KN) | dynamic\_load\_rating | Decimal | kN | 75% | Yes |
| Static Load (kN) | static\_load\_rating | Decimal | kN | 50% | Yes |
| Limiting Speed | limiting\_speed | Decimal | rpm | 37.5% | No |
| Reference Speed | reference\_speed | Decimal | rpm | 22.5% | No |

## 3.3 Classification Attributes (Non-Dimensional)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Current Name | SharedLens Name | Data Type | Fill Rate | Key Attr? |
| Bearing Type | bearing\_type | Enum | 60% | Yes |
| Cage Material | cage\_material | Enum | 67.5% | No |
| Clearance Type | clearance\_type | Enum | 37.5% | No |
| Closure Type | closure\_type | Enum | 65% | Yes |
| Contact Angle | contact\_angle | String | 37.5% | No |
| Number of Rows | number\_of\_rows | Integer | 72.5% | No |

## 3.4 Procurement/Supplier Attributes

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Current Name | SharedLens Name | Data Type | Fill Rate | Notes |
| Manufacturer Name | manufacturer\_name | String | 95% | Not part of engineering spec |
| Manufacturer Part Number | manufacturer\_part\_number | String | 75% | Varies by supplier qualifier |
| Common Supplier Name | common\_supplier\_name | String | 90% | Procurement context only |
| Local Supplier Name | local\_supplier\_name | String | 90% | Procurement context only |

# 4. SharedLens Implementation Recommendations

## 4.1 Data Migration Strategy

**1. Normalize UOM Data**

* • Extract numeric values from text fields ("6069.87LB" → 6069.87)
* • Extract UOM codes from text fields ("6069.87LB" → "lbf")
* • Convert all measurements to canonical SI units (lbf → kN)
* • Store as decimal with storage\_uom = "kN"

**2. Handle Qualifier Field**

* • Current: "Part.SKF", "Part.NTN", "Part.NEW DEPARTURE", "Part.0"
* • Decision: Move supplier-specific data to procurement attributes
* • Keep single canonical engineering spec per part number
* • Supplier variations become procurement package options

**3. Clean Up Data Types**

* • Convert all dimensional strings to decimal
* • Validate number\_of\_rows as integer (currently has 1.0)
* • Standardize date formats

## 4.2 Key Attribute Selection

Recommended key attributes for Ball Bearings category:

* 1. bearing\_type (Enum) - Critical classification
* 2. inside\_diameter (Decimal, mm) - Primary dimension
* 3. outside\_diameter (Decimal, mm) - Primary dimension
* 4. overall\_width (Decimal, mm) - Primary dimension
* 5. dynamic\_load\_rating (Decimal, kN) - Performance spec
* 6. static\_load\_rating (Decimal, kN) - Performance spec
* 7. closure\_type (Enum) - Configuration option

**Rationale:**

These 7 attributes provide sufficient detail to:

* • Identify potential duplicate parts
* • Enable meaningful search (find bearings by size and capacity)
* • Track data completeness (70-100% fill rates)
* • Support engineering decisions

## 4.3 Attribute Aliasing Examples

Example 1: Dynamic Load Rating

* Canonical Storage:
  + dynamic\_load\_rating = 27.00 (storage\_uom: "kN")
* Engineering Lens Alias:
  + Display: "Dynamic Load Rating: 27.0 kN"
  + Transformation: None (canonical units)
* Procurement Lens Alias:
  + Display: "Dynamic Load Rating: 6,069.9 lbf"
  + Transformation: UOM\_Conversion\_SI (kN to lbf, factor 224.809)

Example 2: Inside Diameter

* Canonical Storage:
  + inside\_diameter = 71.975 (storage\_uom: "mm")
* Engineering Lens (Metric):
  + Display: "Bore Diameter: 72.0 mm"
  + Transformation: None
* Engineering Lens (Imperial):
  + Display: "Bore Diameter: 2.834 in"
  + Transformation: UOM\_Conversion\_SI (mm to in, factor 0.0393701)
* Procurement Lens:
  + Display: "Inside Diameter: 2.834 in" (matches supplier catalogs)
  + Transformation: UOM\_Conversion\_SI

# 5. Data Quality Metrics

## 5.1 Current Fill Rates by Category

|  |  |  |  |
| --- | --- | --- | --- |
| Category | Attributes | Avg Fill Rate | Status |
| Core Dimensional | 3 | 100% | Excellent |
| Performance/Load | 4 | 46% | Needs Improvement |
| Classification | 6 | 53% | Acceptable |
| Procurement | 4 | 87% | Good |
| Metadata | 11 | 100% | Excellent |

## 5.2 Key Attribute Completion Goal

**Current State:**

* • 7 recommended key attributes
* • Average completion: ~80%
* • Range: 60% to 100%

**Target State:**

* • Core dimensional (3 attrs): Maintain 100%
* • Performance (2 attrs): Improve from 60% to 85%
* • Classification (2 attrs): Improve from 62% to 90%

**Action Items:**

* 1. Identify 16 records missing dynamic\_load\_rating
* 2. Identify 20 records missing static\_load\_rating
* 3. Identify 16 records missing bearing\_type
* 4. Prioritize data collection for high-volume part numbers

# 6. Implementation Roadmap

## Phase 1: Data Cleanup (2-3 weeks)

* • Task 1.1: Parse and extract UOM data from text fields
* • Task 1.2: Convert all measurements to SI canonical units
* • Task 1.3: Validate data types (decimal vs. string)
* • Task 1.4: Remove unused min/max columns
* • Task 1.5: Standardize enum values (bearing types, materials)

## Phase 2: Schema Definition (1 week)

* • Task 2.1: Define attribute schema per SharedLens model
* • Task 2.2: Mark key attributes for Ball Bearings category
* • Task 2.3: Configure SI UOM relationships
* • Task 2.4: Define validation rules

## Phase 3: Alias Configuration (1-2 weeks)

* • Task 3.1: Define Engineering lens aliases
* • Task 3.2: Define Procurement lens aliases
* • Task 3.3: Configure UOM conversion rules
* • Task 3.4: Test conversions (kN ↔ lbf, mm ↔ in)

## Phase 4: Data Migration (1 week)

* • Task 4.1: Migrate 40 ball bearing records
* • Task 4.2: Validate data quality metrics
* • Task 4.3: Test lens views for correctness
* • Task 4.4: Verify UOM conversions in UI

## Phase 5: Data Quality Improvement (Ongoing)

* • Task 5.1: Target missing key attributes for data collection
* • Task 5.2: Monitor key attribute completion rate
* • Task 5.3: Identify and merge potential duplicates
* • Task 5.4: Establish data governance process

# 7. Conclusion

The Ball Bearings dataset demonstrates typical challenges in attribute management:

* • UOM data embedded in values rather than properly normalized
* • Mixed engineering and procurement data without clear separation
* • Good dimensional data quality (100%) but gaps in performance specs
* • Supplier-specific variations creating multiple records per part

SharedLens implementation will:

* 1. Normalize all UOM data using SI system
* 2. Separate engineering specs from supplier options
* 3. Enable natural terminology per lens (bore vs. inside diameter)
* 4. Support automatic conversions (kN ↔ lbf, mm ↔ in)
* 5. Track and improve key attribute completion rates
* 6. Enable effective deduplication and search

End of Analysis

Ball Bearings - SharedLens Attribute Development

October 23, 2025