

## **Oracle® Enterprise Data Quality for Product Data**

Knowledge Studio リファレンス・ガイド

リリース 5.6.2

部品番号 : B66253-01

2011 年 12 月

Oracle Enterprise Data Quality for Product Data Knowledge Studio リファレンス・ガイド, リリース 5.6.2

部品番号 : B66253-01

Copyright © 2001, 2011 Oracle and/or its affiliates. All rights reserved.

Primary Author: Lorna Vallad

This software and related documentation are provided under a license agreement containing restrictions on use and disclosure and are protected by intellectual property laws. Except as expressly permitted in your license agreement or allowed by law, you may not use, copy, reproduce, translate, broadcast, modify, license, transmit, distribute, exhibit, perform, publish, or display any part, in any form, or by any means. Reverse engineering, disassembly, or decompilation of this software, unless required by law for interoperability, is prohibited.

The information contained herein is subject to change without notice and is not warranted to be error-free. If you find any errors, please report them to us in writing.

If this is software or related documentation that is delivered to the U.S. Government or anyone licensing it on behalf of the U.S. Government, the following notice is applicable:

U.S. GOVERNMENT RIGHTS Programs, software, databases, and related documentation and technical data delivered to U.S. Government customers are "commercial computer software" or "commercial technical data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, duplication, disclosure, modification, and adaptation shall be subject to the restrictions and license terms set forth in the applicable Government contract, and, to the extent applicable by the terms of the Government contract, the additional rights set forth in FAR 52.227-19, Commercial Computer Software License (December 2007). Oracle America, Inc., 500 Oracle Parkway, Redwood City, CA 94065.

This software or hardware is developed for general use in a variety of information management applications. It is not developed or intended for use in any inherently dangerous applications, including applications that may create a risk of personal injury. If you use this software or hardware in dangerous applications, then you shall be responsible to take all appropriate fail-safe, backup, redundancy, and other measures to ensure its safe use. Oracle Corporation and its affiliates disclaim any liability for any damages caused by use of this software or hardware in dangerous applications.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group.

This software or hardware and documentation may provide access to or information on content, products, and services from third parties. Oracle Corporation and its affiliates are not responsible for and expressly disclaim all warranties of any kind with respect to third-party content, products, and services. Oracle Corporation and its affiliates will not be responsible for any loss, costs, or damages incurred due to your access to or use of third-party content, products, or services.

---

---

# 目次

はじめに .....	ix
ドキュメントのアクセシビリティについて .....	x
関連ドキュメント .....	x
表記規則 .....	xi
<b>1 概要</b>	
ソフトウェアの起動 .....	1-2
クライアント・ワークスペースの理解 .....	1-4
フレーム機能 .....	1-4
メニュー・コマンドおよびツールバー .....	1-5
「File」メニュー .....	1-6
「Edit」メニュー .....	1-8
「View」メニュー .....	1-9
「Data Lens」メニュー .....	1-11
「Tools」メニュー .....	1-12
「Help」メニュー .....	1-13
キーボード・ショートカット .....	1-14
タブおよびサブタブ .....	1-14
タスク・ペイン .....	1-15
コンテキスト依存メニュー .....	1-15
Knowledge Studio の起動 .....	1-16
データ・レンズの作成またはオープン .....	1-17
<b>2 データ内のフレーズ</b>	
「Phrases」タブ .....	2-2
「Define Phrases」サブタブ .....	2-2
「Phrases and Terms」ペイン .....	2-2
「Sample Data Table」ペイン .....	2-4
「Item Definition」フィールド .....	2-5
「Graphical Rule Builder」ペイン .....	2-5
「Selected Line」フィールド .....	2-6

ツールの使用 .....	2-7
不明品目に対する用語ルールの作成 .....	2-7
用語ノードに対するフレーズ構造の作成 .....	2-8
2つのフレーズ・ルールを1つのフレーズ・ルールに結合 .....	2-8
2つのフレーズ・ルールのマージ .....	2-9
フレーズ・ルールの名前変更 .....	2-9
フレーズ・ルールの編集 .....	2-10
用語のバリエーション形式の追加 .....	2-10
フレーズの生成結果の変更 .....	2-10
ルール予測の使用 .....	2-11
フレーズのあいまい表現の解決 .....	2-13
用語の完全形式の編集 .....	2-13
親ノードの作成、挿入または切断 .....	2-13
<b>「Define Items」 サブタブ</b> .....	2-14
「Item Definitions」 ペイン .....	2-14
「Sample Data Table」 ペイン .....	2-15
「Item Definition」 フィールドおよび「Selected Line」 フィールド .....	2-15
「Graphical Rule Builder」 ペイン .....	2-16
ツールの使用 .....	2-16
新規の最上位レベル品目定義の作成 .....	2-16
サブ品目定義の作成 .....	2-16
別名 .....	2-17
別名を使用した品目定義および属性のラベル付け .....	2-17
別名の制限 .....	2-17
別名の編集 .....	2-17
別名と属性名が異なる場合の優先ルール .....	2-17
別名情報の表示 .....	2-17
アクティブまたは非アクティブな品目定義 .....	2-18
非アクティブ・フラグの設定 .....	2-18
非アクティブな品目定義の出力データの表示 .....	2-19
「Item Definition List」 ペイン .....	2-20
ツールの使用 .....	2-22
品目定義属性の定義 .....	2-22
属性の追加 .....	2-23
品目定義間での属性のコピーとペースト .....	2-23
属性値ロジックの変更 .....	2-24
属性検索ロジックの変更 .....	2-26
<b>「View Hierarchy」 サブタブ</b> .....	2-27
「Top Down Grammar Hierarchy」 および「Bottom Up Grammar Hierarchy」 ペイン .....	2-27
<b>「Regression Testing」 サブタブ</b> .....	2-28
回帰テストがアクティブであることの確認 .....	2-28
「Before」 ペイン .....	2-29
「After」 ペイン .....	2-29
「Review」 列 .....	2-29
回帰ベースの作成と更新 .....	2-29

### 3 データの標準化

「Standardize」タブ .....	3-2
「Standardize Terms」サブタブ .....	3-2
「Terminology Rules」ペイン .....	3-2
「Rewrite Rules」ペイン .....	3-3
「Copy Full Form」ボタン .....	3-3
「Default Case」セクション .....	3-4
「Standardize Phrases」サブタブ .....	3-5
「Nodes to Receive Rules」ペイン .....	3-5
Sort Productions .....	3-5
「Nodes for Insertion」ペイン .....	3-5
「Ordering Rule」ペイン .....	3-6
用語の結合 .....	3-6
用語の削除または順序変更 .....	3-7
「Standardize Lines」サブタブ .....	3-7
「Unit Conversion」サブタブ .....	3-8
「Nodes to Receive Rules」ペイン .....	3-8
「Instructions」ペイン .....	3-8
「Unit Conversion Table Construction Wizard」ペイン .....	3-8
「Test Global Standardization」サブタブ .....	3-9
サンプル・データ表 .....	3-9
サブタブのデータ・フィールド .....	3-10
「Regression Test」サブタブ .....	3-10
「Before」ペイン .....	3-11
「After」ペイン .....	3-11
「Review」列 .....	3-11
回帰ベースの作成と更新 .....	3-11
標準化のタイプ .....	3-12
標準化タイプの作成 .....	3-12
標準化タイプの削除 .....	3-13

### 4 品目定義の標準化

「Standardize Items」タブ .....	4-2
「Standardize Attributes」サブタブ .....	4-2
「Item Definitions」ペイン .....	4-2
「Rewrite Rules」ペイン .....	4-3
数値リライト・ルール .....	4-3
ボタン .....	4-3
オプション .....	4-4
テキスト置換表 .....	4-4
値置換表 .....	4-4
簡単なリライト・ルール .....	4-5
自動のリライト・ルール .....	4-5
フレーズの生成結果の順序変更 .....	4-5

「Order Attributes」サブタブ .....	4-6
「Item Definition」ペイン .....	4-6
「Selection」ペイン .....	4-6
Null Handling .....	4-6
Multiple Instance Handling .....	4-7
「Match Weights」サブタブ .....	4-8
「Item Definition」ペイン .....	4-8
「Selection」ペイン .....	4-8
重複照合 .....	4-8
「Test Attributes」サブタブ .....	4-9
「Item Definition」セクション .....	4-9
サンプル・データ表 .....	4-10
ソース・フィールド .....	4-10
「Standardized Text」セクション .....	4-10
標準化属性表 .....	4-10
「Test Item Standardization」サブタブ .....	4-11
サンプル・データ表 .....	4-11
ソース・フィールド .....	4-11
「Standardized Text」セクション .....	4-12
「Item Attributes」セクション .....	4-12
「Regression Test」サブタブ .....	4-12
サンプル・データ表 .....	4-13
「Current Attribute」セクションおよび「Regression Attribute」セクション .....	4-13
回帰ベースの作成と更新 .....	4-13
Match Type .....	4-14

## 5 データの分類

分類スキーマ .....	5-2
UNSPSC .....	5-2
eCl@ss .....	5-3
ユーザー定義 .....	5-3
分類タイプ .....	5-3
分類タイプの作成 .....	5-3
分類タイプの削除 .....	5-4
「Classify」タブ .....	5-5
「Classify from Data」サブタブ .....	5-6
「Classification Type」ペイン .....	5-6
「Sample Data」ペイン .....	5-6
「Item Definition」セクション .....	5-7
「Graphical Rule」ペイン .....	5-7
「Classify from Item Definitions」サブタブ .....	5-8
「Classify from Rules」サブタブ .....	5-9
「Test Classification」サブタブ .....	5-10
「Sample Data」ペイン .....	5-10
ソース・フィールド .....	5-10
「Standardized Text」フィールド .....	5-10

「Classification」 ペイン .....	5-10
複数分類のテスト .....	5-10
「Regression Test」 サブタブ .....	5-11
回帰ベースの作成と更新 .....	5-11

## 6 データの翻訳

翻訳プロセス .....	6-2
データ・レンズ標準化の品質 .....	6-2
翻訳に備えたデータの準備 .....	6-2
翻訳ターゲット .....	6-3
翻訳スマート用語集 .....	6-4
Oracle DataLens Server の翻訳スマート用語集の作成 / 更新 .....	6-4
「Translation」 タブ .....	6-5
「New Phrases」 サブタブおよび「Known Phrases」 サブタブ .....	6-5
「Translation」 ペイン .....	6-5
赤色のチェック・マーク .....	6-5
Export .....	6-6
翻訳のためのテキストのエクスポート .....	6-6
翻訳のためのテキストのインポート .....	6-7
「Source Text」 フィールドおよび「Translated Text」 フィールド .....	6-7
「New Variable Term Phrases」 サブタブおよび「Known Variable Term Phrases」 サブタブ .....	6-7
「Reorder」 サブタブ .....	6-8
「Test Translation」 サブタブ .....	6-9
「Test Translated Attributes」 サブタブ .....	6-9
「Item Definition」 セクション .....	6-9
サンプル・データ表 .....	6-10
ソース・フィールド .....	6-10
「Translated Text」 セクション .....	6-10
標準化属性表 .....	6-10
「Test Item Translation」 サブタブ .....	6-11
サンプル・データ表 .....	6-11
ソース・フィールド .....	6-11
「Item Definition」 フィールド .....	6-11
「Test Global Translation」 サブタブ .....	6-12
サンプル・データ表 .....	6-12
ソース・フィールド .....	6-12
「Translated」 フィールド .....	6-12
Quality Metrics .....	6-12
「Regression Test」 サブタブ .....	6-13
「Before」 ペイン .....	6-13
「After」 ペイン .....	6-14
「Review」 列 .....	6-14
回帰ベースの作成と更新 .....	6-14
翻訳修正書式 .....	6-15

## 7 レポートの生成

機能固有のレポート .....	7-2
Classification Reports .....	7-2
Standardization Report .....	7-3
Grammar Reports .....	7-4
Complexity Reports .....	7-5
Semantic Reports .....	7-6

## 8 データ・レンズおよびファイルの管理

データ・レンズの管理 .....	8-2
データ・レンズ・オプションの設定 .....	8-2
データ・レンズのチェックイン .....	8-4
データ・レンズのチェックアウト .....	8-5
データ・レンズの削除 .....	8-5
読取り専用のデータ・レンズの削除 .....	8-6
タスクの表示 .....	8-6
タスク・ステータスの変更 .....	8-7
タスクの作成 .....	8-8
データ・レンズに関する情報 .....	8-9
データ・レンズの摘要の編集 .....	8-9
データ・レンズの履歴ノートの編集 .....	8-9
サンプル・ファイル .....	8-10
新しいサンプル・データの作成 .....	8-10
サンプル・ファイルの削除 .....	8-11
サンプル・ファイルの名前の変更 .....	8-11
サンプル・ファイルの結合 .....	8-12

## 9 エクスポートとインポート

データ・レンズのエクスポート .....	9-2
データ・レンズのインポート .....	9-2
ルールのエクスポート .....	9-2
フレーズおよび用語のインポート .....	9-3
Excel を使用した、ルールおよびフレーズのインポート時の別名の作成 .....	9-4
品目定義のエクスポートとインポート .....	9-4
品目定義のエクスポートとインポートに関するトラブルシューティング .....	9-5
既存の品目定義と同じ名前のインポート品目定義 .....	9-5
スマート用語集に含まれるナレッジに依存するソース品目定義 .....	9-5
標準化情報に依存するソース品目定義 .....	9-5
単位換算情報に依存するソース品目定義 .....	9-5
値ロジック情報に依存するソース品目定義 .....	9-5

## 10 コンテキストおよび品目定義の詳細な定義

属性の編集 .....	10-2
用語バリエーションの生成 .....	10-3
フレーズおよび用語ルールの構文 .....	10-4
用語ルール .....	10-4
用語ルールにおける大 / 小文字の区別 .....	10-4
用語ルールにおける正規表現 .....	10-5
空白、正規表現および用語ルール .....	10-5
フレーズ構造ルール .....	10-6
フレーズ構造ルールにおける開始記号 .....	10-7
ルールでの他の制約 .....	10-7
グローバル・フレーズ・ルールの名前の変更 .....	10-8
ソース書式 .....	10-9
ソース書式設定の適用 .....	10-9
文法のコンパクト設定 .....	10-10

## 11 データの詳細な標準化

用語の標準化 .....	11-2
大 / 小文字の置換 .....	11-2
正規表現の置換 .....	11-3
個別の置換 .....	11-4
複数の標準化タイプ .....	11-4
単位の標準化タイプ .....	11-5
単位換算タイプの作成 .....	11-5
単位換算タイプの削除 .....	11-6
フレーズに対する単位換算の作成 .....	11-6
データ・レンズでの標準化の共有 .....	11-9
グローバル標準化のコピー .....	11-10
品目定義標準化の共有 .....	11-11
標準化タイプ間での品目定義のコピー .....	11-12
品目定義間のコピー .....	11-15

## 12 データの詳細な分類

高度な分類ルールの機能 .....	12-2
追加 .....	12-2
マスキング .....	12-2
否定 .....	12-2
親 .....	12-2
機能の例 .....	12-3
マスクの追加 .....	12-3
否定の追加 .....	12-3
親の追加 .....	12-3
複数の分類スキーマ .....	12-4
分類タイプのアップグレード .....	12-4

ユーザー定義の分類タイプ .....	12-5
Excel を使用したユーザー定義のスキーマの作成 .....	12-5
テキスト・エディタを使用したユーザー定義のスキーマの作成 .....	12-6
ユーザー定義の分類タイプの作成 .....	12-6
グローバル分類スキーマ .....	12-6
分類システムのサポート .....	12-7
データ・レンズ分類システムのファイル形式 .....	12-7

## 13 スマート用語集

スマート用語集のインポート .....	13-2
スマート用語集の作成 .....	13-3
付属のスマート用語集 .....	13-4
色 .....	13-4
数 .....	13-4
販売用パッケージ設定 .....	13-5
単位 .....	13-6
単位（小売） .....	13-7

## A クライアント・ソフトウェアのインストール

## B 正規表現

正規表現で使用する特殊文字 .....	B-2
用語ルールで役立つ正規表現 .....	B-4

---

---

# はじめに

このリファレンス・ガイドでは、Oracle Enterprise Data Quality for Product Data Knowledge Studio の基本的な機能を説明します。このドキュメントの構成は次のとおりです。

- 第1章～第9章では、アプリケーションの基本機能について説明します。
- 第10章～第13章では、拡張機能について説明します。

理解するには、このリファレンス・ガイドとともに、x ページの「[関連ドキュメント](#)」に掲載されている Oracle Enterprise Data Quality for Product Data の各ドキュメントも併せてお読みください。

このマニュアルを使用する前に、次の Oracle Enterprise Data Quality for Product Data ドキュメントを確認してください。

- Oracle Enterprise Data Quality for Product Data Application Studio リファレンス・ガイド
- Oracle Enterprise Data Quality for Product Data Governance Studio リファレンス・ガイド

## ドキュメントのアクセシビリティについて

オラクル社のアクセシビリティについての詳細情報は、Oracle Accessibility Program の Web サイト <http://www.oracle.com/pls/topic/lookup?ctx=acc&id=docacc> を参照してください。

### Oracle サポートへのアクセス

Oracle のお客様は、My Oracle Support にアクセスして電子サポートを受けることができます。詳細は、<http://www.oracle.com/pls/topic/lookup?ctx=acc&id=info> または <http://www.oracle.com/pls/topic/lookup?ctx=acc&id=trs> (聴覚障害者向け) を参照してください。

## 関連ドキュメント

詳細は、ドキュメント・セットに含まれる次のドキュメントを参照してください。

- Oracle Enterprise Data Quality for Product Data Oracle DataLens Server インストールेशन・ガイドには、Oracle DataLens Server のインストール手順の詳細が記載されています。
- Oracle Enterprise Data Quality for Product Data Oracle DataLens Server 管理者ガイドには、Oracle DataLens Server のインストールと管理に関する情報が記載されています。
- Oracle Enterprise Data Quality for Product Data COM インタフェース・ガイドには、Oracle DataLens Server の COM API のインストールと使用に関する情報が記載されています。
- Oracle Enterprise Data Quality for Product Data Java インタフェース・ガイドには、Oracle DataLens Server の Java API のインストールと使用に関する情報が記載されています。
- Oracle Enterprise Data Quality for Product Data Application Studio リファレンス・ガイドには、Data Service Applications (DSA) の作成と保守に関する情報が記載されています。
- Oracle Enterprise Data Quality for Product Data AutoBuild リファレンス・ガイドには、既存の製品情報およびデータ・レンズ・ナレッジに基づいた初期データ・レンズの作成に関する情報が記載されています。
- Oracle Enterprise Data Quality for Product Data Governance Studio リファレンス・ガイドには、Data Service Applications (DSA) の実行と出力データの分析に関する情報が記載されています。
- Oracle Enterprise Data Quality for Product Data 用語集には、Enterprise DQ for Product テクノロジーで一般的に使用する用語の定義が記載されています。
- Oracle Enterprise Data Quality for Product Data Services for Excel リファレンス・ガイドには、Microsoft Excel スプレッドシートに含まれるデータに基づいた DSA の作成に関する情報が記載されています。
- Oracle Enterprise Data Quality for Product Data Task Manager リファレンス・ガイドには、Task Manager または Governance Studio アプリケーションで作成されたタスクの管理に関する情報が記載されています。

このドキュメントおよびすべての関連ドキュメントの最新版は、次に示す Oracle Enterprise Data Quality for Product Data ドキュメントの Web サイトから入手できます。

[http://download.oracle.com/docs/cd/E20593\\_01/index.htm](http://download.oracle.com/docs/cd/E20593_01/index.htm)

## 表記規則

このマニュアルでは、次の表記規則を使用します。

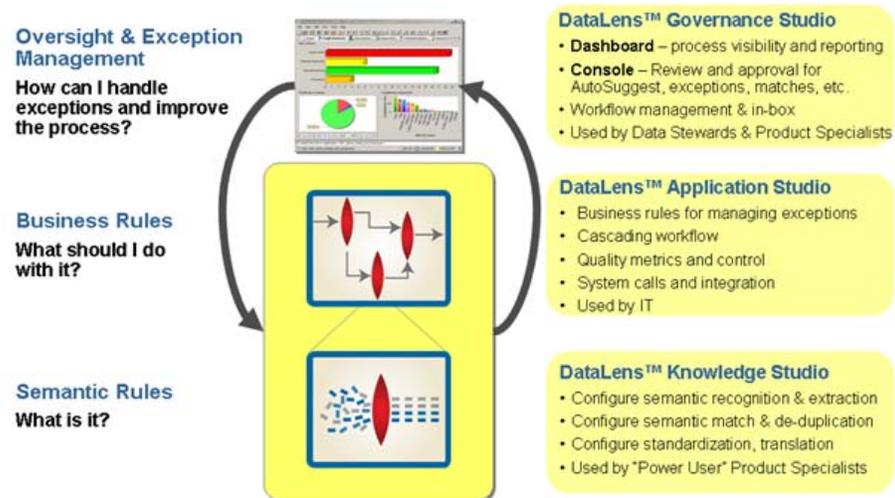
規則	意味
太字	太字は、操作に関連する Graphical User Interface 要素、または本文中で定義されている用語および用語集に記載されている用語を示します。
イタリック体	イタリックは、マニュアルのタイトル、強調、またはユーザーが特定の値を指定するプレースホルダ変数を示します。
固定幅フォント	固定幅フォントは、段落内のコマンド、URL、サンプル内のコード、画面に表示されるテキスト、入力するテキスト、ファイル、ディレクトリまたはパス名を示します。
固定幅フォント	太字の固定幅フォントは、入力するコマンドまたはテキストを示します。



## 概要

Oracle DataLens Server は、様々なソースおよびシステムの製品データを標準化、照合、補完および修正する業界最先端の DataLens™ テクノロジーを基に構築されています。DataLens の中核テクノロジーには、製品データに特有の変動性の高さに対応するために、一から設計された特許取得済セマンティック・テクノロジーが使用されています。

Oracle Enterprise Data Quality for Product Data（以前の Oracle Product Data Quality）では、DataLens の中核テクノロジー・モジュールである Governance Studio、Knowledge Studio および Application Studio が使用されます。次の図は、これらのモジュールのプロセス・フローを示しています。



このガイドでは、Enterprise DQ for Product (EDQP) Knowledge Studio の効果を最大限にするために使用できる基本的なテクニックと高度なテクニックについて説明します。これらのテクニックを使用すると、データに関する知識をより正確にして、業務の専門家 (SME) に対して、DataLens 手法の重要な側面に関する詳細な情報を簡単に提供できます。

Knowledge Studio を使用すると、データの認識、分類および標準化を可能にするルールのコレクションである、データ・レンズを作成できます。データ・レンズの作成に必要な主要アクティビティは3つあります。

### データの認識：

データを認識するルールを作成し、バリエーション形式をレンズに組み込みます。

### 品目の定義：

品目を正確に定義するために必要な属性を識別します。

**データの標準化：**

用語、フレーズおよび品目定義に関する標準化ルールを作成します。

このリファレンス・ガイドでは、筆記用具の製品データを使用して、データ・レンズの作成プロセスを理解できるようにします。

## ソフトウェアの起動

Enterprise DQ for Product を起動するには、次のように、デスクトップ・ショートカット、または Windows の「スタート」メニューを使用します。

---

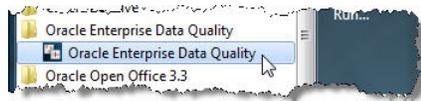
**注意：** Enterprise DQ for Product がインストールされていない場合は、[付録 A 「クライアント・ソフトウェアのインストール」](#)の手順を使用してインストールできます。

---

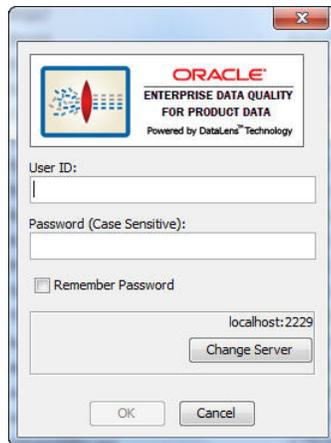
- デスクトップ・ショートカットをダブルクリックします。



- 「スタート」、「プログラム」、「Oracle Enterprise Data Quality for Product Data」の順にクリックし、「Oracle Enterprise Data Quality for Product Data」を選択します。

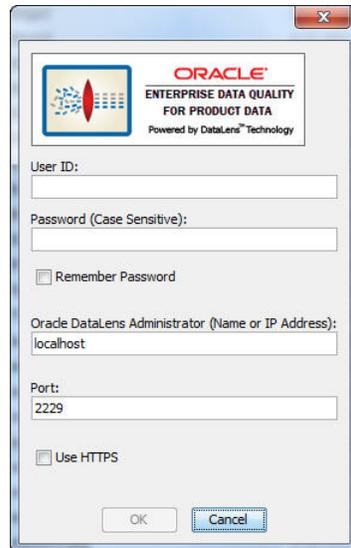


「Oracle Enterprise Data Quality for Product Data Login」ダイアログ・ボックスが表示されます。



ユーザー名とパスワードを入力します。「Remember Password」チェック・ボックスを選択すると、ログインするたびにパスワードを入力せずに済みます。

Oracle DataLens Server を変更するか、HTTP Secure (HTTPS) を使用する場合は、「Change Server」をクリックします。HTTPS オプションは、アプリケーション・サーバーとして WebLogic を使用する Oracle DataLens Server で実行する場合のみ動作保証されています。



Oracle DataLens Server を変更するには、Oracle DataLens Administration Server のホスト名または IP アドレスとそのポート番号を入力します。

---

**注意：** Oracle DataLens Server で WebLogic アプリケーション・サーバーを使用している場合、このポート番号は、WebLogic ドメイン・サーバーのポート番号と一致する必要があります。

---

HTTPS を使用して Oracle DataLens Server に接続するには、「Use HTTPS」チェック・ボックスを選択します。

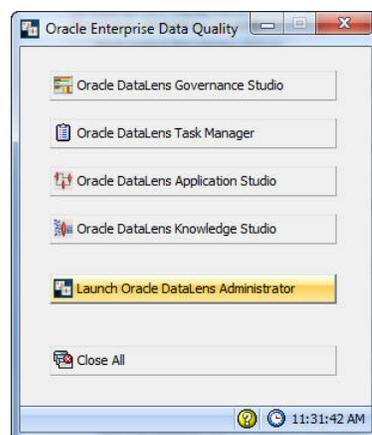
---

**注意：** Oracle DataLens Server（および EDQP トポロジ内の他のすべてのサーバー）はすべて、HTTPS を使用するように構成する必要があります。HTTPS の構成および使用の詳細は、*Oracle Enterprise Data Quality for Product Data Oracle DataLens Server 管理者ガイド*を参照してください。

---

情報がすべて正しい場合は、「OK」をクリックします。

「Oracle Enterprise Data Quality for Product Data Launch Pad」が表示されます。



「Oracle Enterprise Data Quality for Product Data Launch Pad」を使用すると、ボタンをクリックすることによって、いずれかの Oracle DataLens Server アプリケーションをすばやく起動できます。「Close All」ボタンを使用すると、オープンしているすべての Enterprise DQ for Product アプリケーションをクローズできます。

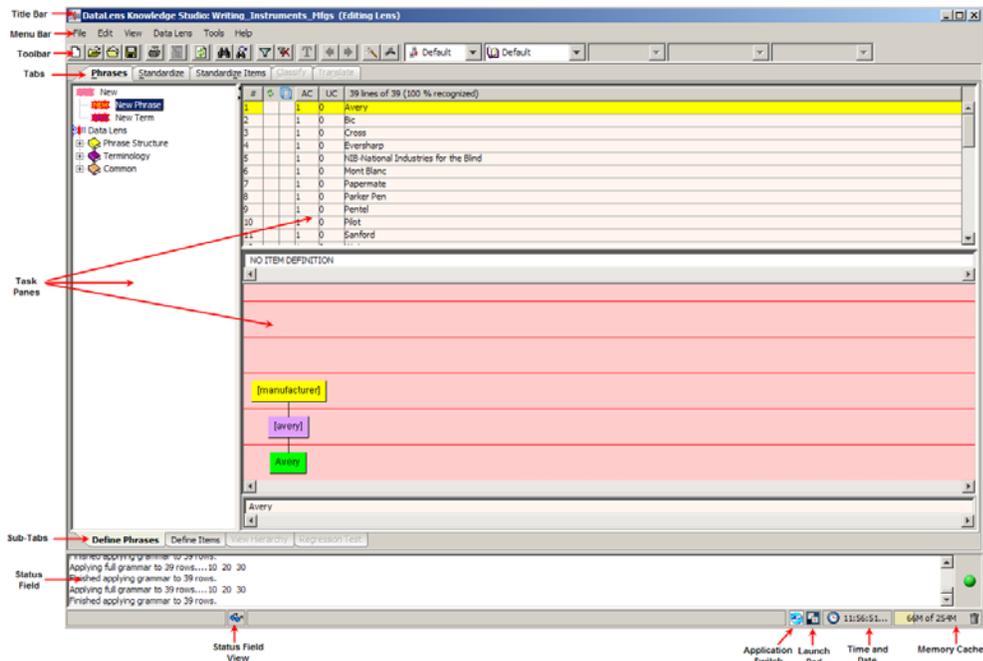
「Oracle Knowledge Studio」ボタンをクリックして、アプリケーションを起動します。

## クライアント・ワークスペースの理解

Knowledge Studio の Graphical User Interface (GUI) では、データ・レンズの作成および管理に使用するクライアント・ワークスペースが提供されます。

- 1-4 ページの「フレーム機能」
- 1-5 ページの「メニュー・コマンドおよびツールバー」
- 1-14 ページの「タブおよびサブタブ」
- 1-15 ページの「タスク・ペイン」

**注意：** 設定されていない機能、または現在のユーザーに使用する権限がない機能は、グレー表示されます。



## フレーム機能

Knowledge Studio のクライアント・ワークスペース・フレームには、次のような有用な情報および対話機能が含まれています。

### タイトル・バー

現在のアプリケーションおよびオープンしているプロジェクトを示します。

### ステータス・フィールド

データ・レンズの処理ステータスが一度に 1 行ずつ表示されます。このフィールドのサイズは変更でき、右側にあるスクロール矢印を使用すると、使用可能なステータス情報をすべて表示できます。ステータス・データは選択したタブに基づいて変わることはなく、すべてのデータを編成したものです。

**ステータス・フィールド表示**

「Status Field」を表示するかどうかを制御します。

**アプリケーション・スイッチ**

前回使用した Enterprise DQ for Product アプリケーションに戻ります。

**Oracle Enterprise Data Quality for Product Data Launch Pad**

このボタンを使用すると、Oracle Enterprise Data Quality for Product Data Launch Pad がオープンし、他のアプリケーションを選択できます。

**時間と日付**

時間が表示され、このフィールドにカーソルをあわせると日付が表示されます。

**メモリー・キャッシュ**

現在使用されているメモリー・キャッシュの量、および使用可能な合計量を示します。この対話フィールドのゴミ箱アイコンをクリックすると、メモリー・キャッシュをダンプできます。

---

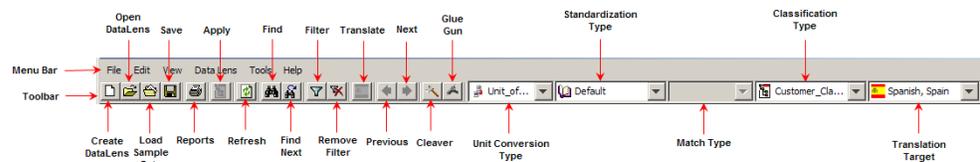
**注意：** この機能は、システム診断用のみ使用され、サポート・チームによって要求されないかぎり使用しないでください。

---

**メニュー・コマンドおよびツールバー**

Knowledge Studio のツールバーを使用すると、最も頻繁に使用する Knowledge Studio の機能に簡単にアクセスできます。ツールバー・ボタンのセットは、ユーザー・インタフェース操作時は同じままですが、インタフェースの現在の状態および設定したオプションによって使用可能または使用不可になります。グレー・スケールで表示されるボタンは使用不可です。フルカラーのボタンは使用可能です。ツールバー・ボタンはすべて標準のプッシュ・ボタンで、マウスを1回クリックするとアクティブ化されます。

次に、ツールバーのボタンを左から右に簡単に説明します。



Knowledge Studio の GUI メニューから、ほとんどの Knowledge Studio の機能にアクセスできます。ツールバーのすべてのボタンには対応するメニュー・コマンドがあり、各メニューでは、コマンドの横に表示されるボタン・アイコンとして示されます。メニュー・コマンドのセットは、GUI 操作時は同じままです。

メニュー・コマンドは、データ・レンズの現在の状態に基づいて使用可能または使用不可になり、グレー表示されたコマンドは使用できません。一部のメニュー・コマンドでは、より複雑な省略記号 (...) で示される機能が実行されます。これらのコマンドでは、要求した機能の完了に必要な情報を収集するダイアログがオープンします。ユーザー機能を切り替えるメニュー・コマンドの前には、チェック・マークが付きます。

**ヒント：** マウス・ポインタをメニュー項目、ボタン、タブ、アイコンまたは同様のコンテンツにあわせると、ツールヒントが表示されます。

次の各項では、各 Knowledge Studio メニュー・コマンドと対応するボタンについて簡単に説明します。

## 「File」メニュー

### New Data Lens...

データを処理するための新しいデータ・レンズ・ファイルを作成します。データ・レンズ・ファイルは、次のディレクトリに格納されます。

C:\¥Documents and Settings¥Username¥Applications¥DataLens¥data¥project

または

C:\¥Users¥Username¥AppData¥DataLens¥export

### Open Data Lens...

既存のデータ・レンズ・ファイルをオープンし、オープンしているデータ・レンズ・ファイルをクローズします。

### Recent Lens

データ・レンズをすばやくオープンできるように、最近オープンしたデータ・レンズのリストが選択用に表示されます。

### Select Data File

現在のデータ・レンズに関連付けられたサンプル・データ・ファイルをオープンし、現在オープンしているサンプル・データ・ファイルをクローズします。

### Close Data Lens

オープンしているデータ・レンズ・ファイルをクローズします。

### Save

すべてのコンテキスト変更をディスクに保存し、元に戻すことができるデータ・レンズのバージョンを作成します。

### Save As

現在のデータ・レンズを新しい名前でも保存できます。

### Delete Data Lens

オープンしているデータ・レンズをローカル・マシンから削除できます。削除の前に警告メッセージが表示されます。データ・レンズのローカル・コピーのみが削除されます。データ・レンズをサーバーにチェックインした場合、そのコピーはサーバー上にまだ存在し、サーバーから削除する必要があります。詳細は、8-5 ページの「[データ・レンズの削除](#)」を参照してください。

### Delete Read-Only Lenses...

不要な読取り専用データ・レンズをローカル・マシンから削除できます。詳細は、8-6 ページの「[読取り専用のデータ・レンズの削除](#)」を参照してください。

### Delete Sample Files...

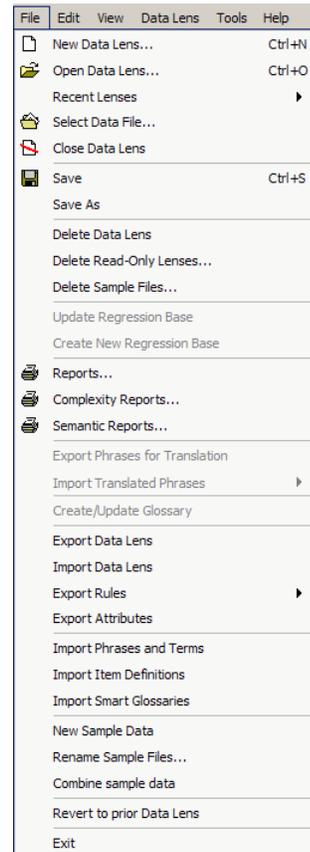
現在編集中のデータ・レンズに関連付けられたサンプル・ファイルを削除できます。削除の対象として「All」を指定するか、または特定のサンプル・ファイルを指定できます。詳細は、8-11 ページの「[サンプル・ファイルの削除](#)」を参照してください。

### Update Regression Base

現在オープンしているタブのコンテキスト変更に基づいて、現在の回帰テスト・ベースを更新できます。

### Create New Regression Base

新しい回帰ベース・ファイルを作成し、このファイルによって、コンテキスト変更が用語およびフレーズに加えられた際の変更の影響を識別します。



---

## Reports

結果を表示するレポート書式を選択できます。詳細は、7-5 ページの「[Complexity Reports](#)」を参照してください。

### Complexity Reports

データの複雑度を示すレポートを選択できます。詳細は、7-5 ページの「[Complexity Reports](#)」を参照してください。

### Semantic Reports

選択したデータ・レンズ内のデータの解析済フレーズ・コンテキストをカウントするレポートを選択できます。詳細は、7-6 ページの「[Semantic Reports](#)」を参照してください。

### Export Phrases for Translation

翻訳ディクショナリからフレーズをエクスポートします。詳細は、6-5 ページの「[Translation](#) タブ」を参照してください。

### Import Translated Phrases

翻訳ディクショナリにフレーズをインポートします。

### Import Current/All Translated Phrases

フレーズの一部またはすべてをインポートします。詳細は、6-5 ページの「[Translation](#) タブ」を参照してください。

### Create/Update Glossary

Oracle DataLens Server 上の用語集を作成または更新できます。詳細は、13-3 ページの「[スマート用語集の作成](#)」を参照してください。

### Export Data Lens

データ・レンズ・プロジェクトから完全な結果およびデータ・レンズをエクスポートし、データ・レンズ・エクスポート・ファイル・プロジェクト・ディレクトリを作成します。

```
¥Documents and Settings¥Username¥Application Data¥DataLens¥export¥data lens name
```

詳細は、9-2 ページの「[データ・レンズのエクスポート](#)」を参照してください。

### Import Data Lens

指定したエクスポート・ディレクトリからエクスポート済データ・レンズをインポートします。詳細は、9-2 ページの「[データ・レンズのインポート](#)」を参照してください。

### Export Rules

用語およびフレーズのルールをエクスポートできます。詳細は、9-2 ページの「[ルールのエクスポート](#)」を参照してください。

### Export Rules by Domain

用語およびフレーズのルールをドメイン別にエクスポートできます。詳細は、9-2 ページの「[ルールのエクスポート](#)」を参照してください。

### Export Attributes

属性を（品目定義から）Excel スプレッドシート・ファイルにエクスポートできます。レポートには、属性タイプ、属性別名、属性名、各標準化に対する属性と順序を定義するルールを示す属性情報が品目定義レベルで表示されます。

### Import Phrases and Terms

ナレッジ（主に用語ルール）を、Excel スプレッドシートまたはタブ区切りファイルからデータ・レンズにインポートできます。詳細は、9-3 ページの「[フレーズおよび用語のインポート](#)」を参照してください。

### Import Item Definitions

品目定義を、タブ区切りファイルからデータ・レンズにインポートできます。詳細は、9-4 ページの「[品目定義のエクスポートとインポート](#)」を参照してください。

---

### Import Smart Glossaries

現在のデータ・レンズのフレーズ編集を容易にするために、基盤データ・レンズをインポートできます。詳細は、13-2 ページの「スマート用語集のインポート」を参照してください。

### New Sample Data

新しいサンプル・データ・ファイルを作成して、既存のサンプル・セットに追加できます。詳細は、8-10 ページの「サンプル・ファイル」を参照してください。

### Rename Sample Files...

データ・レンズに関連付けられた既存のサンプル・ファイルの名前を変更できます。詳細は、8-10 ページの「サンプル・ファイル」を参照してください。

### Combine sample data

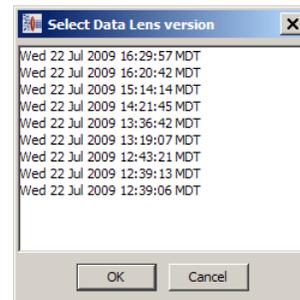
選択したサンプル・ファイルを、回帰テストに使用するための単一ファイルに結合できます。詳細は、8-10 ページの「サンプル・ファイル」を参照してください。

### Revert to prior Data Lens

現在のデータ・レンズの以前のバージョンに戻ることができます。リストされるデータ・レンズはローカル・コピーのみであり、Oracle DataLens Server 内のデータ・レンズはリストされません。

### Exit

Knowledge Studio アプリケーションを終了します。保存していない変更がある場合は、プロンプトが表示されます。



## 「Edit」メニュー

### Cut

選択範囲を削除し、クリップボードにコピーします。

### Copy

選択範囲をクリップボードにコピーします。

### Paste

クリップボードの内容を現在の挿入ポイントにペーストします。

### Replace

「Translate」タブで指定したテキストを検索し、置換します。

### Rename Rules

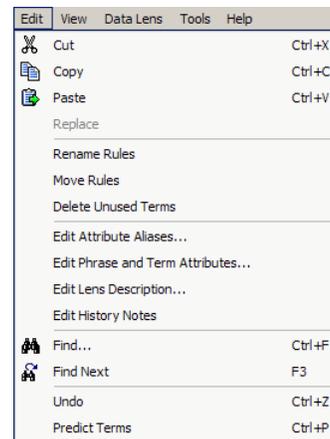
フレーズ・ルールの名前をグローバルに変更し、フレーズ・ルールを統合できます。この機能は、「Phrases」タブの「Define Phrase」サブタブでのみ使用できます。詳細は、10-8 ページの「グローバル・フレーズ・ルールの名前の変更」を参照してください。

### Move Rules

階層フォルダ形式の「Move Rules」ダイアログで、ドメイン（フォルダ）間でルールをドラッグ・アンド・ドロップできます。たとえば、スマート用語集からデータ・レンズのフレーズ構造にルールを移動できます。

### Delete Unused Terms

未使用の用語を削除できます。未使用の用語とは、ルールまたはフレーズによって参照されていない用語です。これは、紫色の丸いアイコンに「u」の文字で示されます。



**Edit Attributes Aliases...**

フレーズおよび用語の属性別名を編集できます。詳細は、2-17 ページの「**別名**」を参照してください。

**Edit Phrase and Term Attributes...**

フレーズおよび用語の属性を編集できます。詳細は、10-2 ページの「**属性の編集**」を参照してください。

**Edit Lens Description...**

データ・レンズの概要を変更できます。詳細は、8-9 ページの「**データ・レンズの概要の編集**」を参照してください。

**Edit History Notes**

継続サポート用の監査証跡を提供するために、データ・レンズの保守に関するテキストを入力できます。基盤またはドメインがデータ・レンズにインポートされる場合は、この情報が日付およびタイムスタンプとともに含まれます。詳細は、8-9 ページの「**データ・レンズの履歴ノートの編集**」を参照してください。

**Find...**

検索文字列（正規表現）を指定して、検索を実行できます。Knowledge Studio の作成タブ（「**Phrases**」、「**Standardize**」および「**Classify**」タブ）の左側のツリー・ペインが検索されます。

**Find Next**

「**Find**」操作で定義した前回の検索を繰り返します。

**Undo**

データ・レンズに加えた変更を削除し、前回保存した状態に戻します。

**Predict Terms**

予測しきい値を満たした信頼度に基づいて、各サンプル行の入力データに適用可能なすべてのルールが表示され、ルールを選択できます。予測が表示されない場合は、その事由を示すメッセージが表示されます。「**Predict Terms**」は、サンプル行に品目定義が関連付けられている場合に、品目定義のコンテキストでのみ機能します。適切なルールを選択するか、または予測を拒否できます。予測の拒否は、現在のデータ・レンズ編集セッションに対してのみ適用可能で、データ・レンズをクローズするとリセットされます。

**「View」メニュー****View My Tasks**

スケジュールしたタスクまたは実行したタスクを表示できます。詳細は、8-6 ページの「**タスクの表示**」を参照してください。

**Filter...**

表示されたデータをテキストまたはテキスト・パターンに基づいてフィルタ処理できます。フィルタ処理操作は、現在選択しているタブにのみ適用されます。「**Filter**」ダイアログに入力したテキストに一致する行のみがタスク・ペインに表示されます。

**Remove Filter**

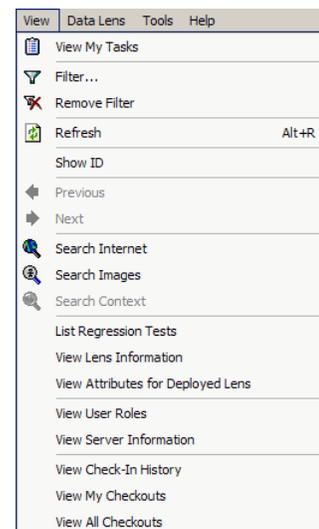
現在適用されているフィルタが削除され、すべてのデータが表示されます。

**Refresh**

「**Apply**」機能を使用して適用した変更を含めて、データが再表示されます。

**Show ID**

選択すると表形式のペインに ID 列が表示され、再度選択すると列がタスク・ペインから削除されます。



---

### Next

次のフレーズまたはルールあいまい表現に進みます。

### Previous

前のフレーズまたはルールあいまい表現に戻ります。

### Search Internet

「Phrases」タブの「Define Phrases」または「Define Items」サブタブにある「Input Data」フィールドで選択したテキストを、インターネットで検索できます。デフォルトのブラウザ・アプリケーションが起動され、選択したテキストを検索文字列として使用して検索が実行されます。

### Search Images

「Phrases」タブの「Define Phrases」または「Define Items」サブタブにある「Input Data」フィールドで選択したテキストに一致するイメージを、インターネットで検索できます。デフォルトのブラウザ・アプリケーションが起動され、選択したテキストを検索文字列として使用してイメージ検索が実行されます。

### Search Context

別のコンテキストで選択できるように、データの選択した行を検索できます。この機能は、「Translation」タブでのみ使用できます。

### List Regression Tests

選択したデータ・レンズに関連付けられた回帰テストに関する情報が表示されます。作成された回帰のタイプ、および回帰テストの対象となるサンプル・ファイルが表示されます。

### View Lens Information

現在使用されているデータ・レンズおよびデータ・ファイルに関する特定の情報が表示されます。

### View Attributes for Deployed Lens

属性の使用を含め、現在デプロイ済のデータ・レンズに関する属性情報が品目定義別に表示されます。

### View User Roles

現在のユーザー・ロール情報が表示されます。

### View Server Information

Oracle DataLens Server に関するサーバー情報が表示されます。

### View Check-In History

チェックインに関するコメントを含め、チェックインしたデータ・レンズがリストされます。

### View My Checkouts

チェックアウトしたデータ・レンズがリストされます。

### View All Checkouts

チェックアウトされた Oracle DataLens Server 上のデータ・レンズがすべてリストされます。

### View as Production

非アクティブに設定された品目定義から出力データが表示されます。詳細は、2-18 ページの「[アクティブまたは非アクティブな品目定義](#)」を参照してください。

---

## 「Data Lens」メニュー

### Check-In Data Lens...

データ・レンズ・ファイルを Oracle DataLens Server リポジトリにチェックインできます。データ・レンズを Oracle DataLens Server にチェックインするたびに、データ・レンズ・リビジョン番号が増加します。Oracle DataLens Server では、データ・レンズの以前のリビジョンがすべて保守されず、データ・レンズをチェックインするには、以前に一度もチェックインされていない、または以前にチェックアウトされていて自分で編集するためにロック済であるという2つの条件のいずれかを満たす必要があります。「**Check-In**」ダイアログを使用すると、データ・レンズのこのリビジョンとともに格納されるコメントを入力できます。データ・レンズの編集を続行する場合、「**Keep Locked for More Editing**」チェック・ボックスを選択すると、別のユーザーはそのデータ・レンズを読み取り専用モードでしかチェックアウトできなくなります。このオプションを選択すると、データ・レンズのローカル・コピーをクライアントから削除する「**Delete local Data Lens**」コマンドはグレー表示になります。詳細は、8-4 ページの「[データ・レンズのチェックイン](#)」を参照してください。

### Check-Out Data Lens...

Oracle DataLens Server リポジトリからチェックアウトするデータ・レンズおよび特定のリビジョン番号を選択でき、データ・レンズは編集用に自動的にロックされます。データ・レンズをチェックアウトして新しい名前を割り当てると、既存のデータ・レンズから新しいデータ・レンズを作成することもできます。詳細は、8-5 ページの「[データ・レンズのチェックアウト](#)」を参照してください。

### Unlock Data Lens

Oracle DataLens Server 内のリポジトリから、現在のデータ・レンズのロックを解除します。

### Check-In Component...

この機能は非推奨です。

### Check-Out Component...

この機能は非推奨です。

### Unlock Component

この機能は非推奨です。

### Copy Global Standardizations

現在の標準化タイプから別の標準化タイプにグローバル標準化ルールをコピーします。詳細は、11-10 ページの「[グローバル標準化のコピー](#)」を参照してください。

### Apply

作成したナレッジをアクティブ化します。変更の適用後、サンプル・データへの影響を確認するには、「**Refresh**」コマンドを使用します。

### Translate

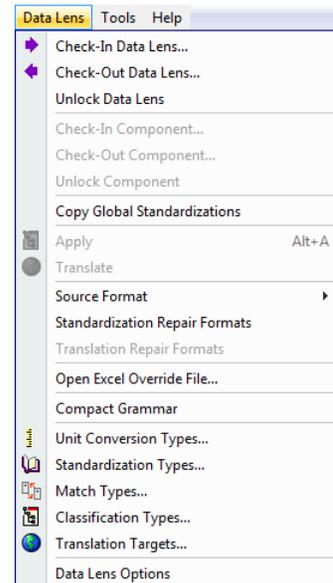
フレーズ（「**Translate**」タブ）、またはコンテンツ行全体（「**Test Translations**」タブ）の翻訳を実行します。詳細は、6-1 ページの「[データの翻訳](#)」を参照してください。

### Source Format

ソースの書式設定を編集できます。詳細は、10-9 ページの「[ソース書式](#)」を参照してください。

### Standardization Repair Formats

標準化したデータを修正するための sed スクリプトを入力できます。



### Translation Repair Formats

翻訳したデータを修正するための sed スクリプトを入力できます。詳細は、6-15 ページの「[翻訳修正書式](#)」を参照してください。

### Open Excel Override File

このデータ・レンズ内で使用される特定のコンテキストの入力に使用できるスプレッドシートを表示して Excel を起動します。この機能は非推奨になる予定であるため使用しないでください。

### Compact Grammar

レンズ内のデータに基づいて、利用されていない文法ルールを削除できます。詳細は、10-10 ページの「[文法のコンパクト設定](#)」を参照してください。

### Unit Conversion Types...

Enterprise DQ for Product 提供の単位換算を追加、選択およびアクティブ化できます。単位換算を使用すると、一貫した単位の使用による出力の作成が可能になります。たとえば、データが抵抗をオーム、キロオームおよびメガオームで表しているとします。単位換算を使用して前述の各単位をオームに換算することで、出力の一貫性を維持できます。詳細は、11-5 ページの「[単位の標準化タイプ](#)」を参照してください。

### Standardization Types...

Enterprise DQ for Product 提供の単位換算を追加、選択およびアクティブ化できます。標準化タイプでも、データ・レンズ全体で使用するための独自の標準化スキーマを作成できます。詳細は、3-12 ページの「[標準化のタイプ](#)」を参照してください。

### Match Types...

データを自動的に照合するためのスキーマを追加および使用できます。詳細は、4-14 ページの「[Match Type](#)」を参照してください。

### Classification Types...

データを自動的に分類するためのスキーマを追加および使用できます。詳細は、5-3 ページの「[分類タイプ](#)」を参照してください。

### Translation Targets...

データ翻訳対象のロケール / 言語を選択できます。このオプションは、データ・レンズが標準化されるまでは使用できません。「**Translation**」タブをアクティブ化します。詳細は、6-3 ページの「[翻訳ターゲット](#)」を参照してください。

### Data Lens Options

テキストの大 / 小文字の区別、データ・レンズをインポートできるかどうか、および「**Apply**」機能の動作など、グローバル・データ・レンズ・パラメータを選択できます。詳細は、8-2 ページの「[データ・レンズ・オプションの設定](#)」を参照してください。

## 「Tools」メニュー

### Open Oracle DataLens Governance Studio...

EDQP Governance Studio を起動します。Oracle Enterprise Data Quality for Product Data Governance Studio リファレンス・ガイドを参照してください。

### Open Oracle DataLens Task Manager...

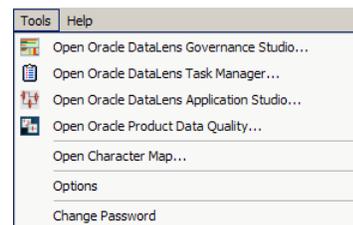
EDQP Task Manager を起動します。Oracle Enterprise Data Quality for Product Data Task Manager ガイドを参照してください。

### Open Oracle DataLens Application Studio...

EDQP Application Studio を起動します。Oracle Enterprise Data Quality for Product Data Application Studio リファレンス・ガイドを参照してください。

### Open Oracle Enterprise Data Quality for Product Data...

Oracle Enterprise Data Quality for Product Data Launch Pad を起動します。



## Open Character Map...

「Windows Character Map」ダイアログを開き、文字マッピングの変更を可能にします。この機能は、フレーズの翻訳時にキーボードで使用できない特殊文字および記号を挿入する簡単な方法として用意されています。

## Options

「Options」ダイアログから次のアプリケーション・オプションを選択できます。

### Parse Tree Node Font Size

「Define Phrases」タブの「Graphical Rule Editor」ペインにあるフレーズ・ツリーの表示に使用するフォント・サイズを選択できます。小さいフォントを使用すると、長い行でより多くのフレーズを表示できます。

### Number of Apply's before Save

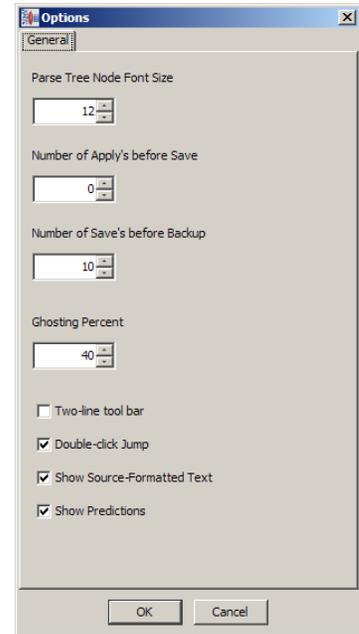
ナレッジの適用時に、データ・レンズを自動的に保存できます。

### Number of Saves before Backup

決定した回数の保存後、データ・レンズを自動的にバックアップできます。

### Ghosting Percent

品目定義に関連付けられていない用語およびフレーズの表示に使用される不明瞭化の割合を設定できます。割合は10～100%にできます。低い割合を設定すると、用語およびフレーズはより薄く（より不明瞭に）なります。



### Two-line tool bar

ツールバーを1行に表示するか2行に表示するかを選択できます。2行のツールバーを選択すると、Knowledge Studio 画面が通常のサイズより小さい場合でも、すべてのツールバー項目を表示できます。

### Double-click Jump

ペインの空白の領域をダブルクリックすることによって、選択したノードの表示をジャンプ（切り替えることが）できます。この機能はコンテキスト依存であり、アクティブなタブを変更します。

### Show Source-Formatted Text

ソース書式設定機能によって書式が再設定されたテキストの表示を可能にして、さらに標準化するためにこのデータをすばやく識別できるようにします。

### Show Predictions

不明なデータ・ノードに対する予測のテキスト表示を可能にします。「Edit」メニューの予測オプション、および「Define Phrases」サブタブの「Graphical Rule Builder」ペインのコンテキスト依存メニューに表示される予測オプションがアクティブかどうかを制御します。

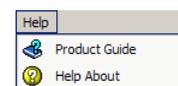
### Change Password

パスワードを変更できます。このチェック・ボックスを選択すると「Change Password」ダイアログが起動して、古いパスワードに続いて新しいパスワードを入力できます。

## 「Help」メニュー

### Product Guide

選択用の Enterprise DQ for Product ドキュメント・リストをブラウザでオープンします。



### Help About

バージョン番号などの製品に関する情報や、サード・パーティ製品のライセンスを表示するリンクが表示されます。

## キーボード・ショートカット

次の表に、Knowledge Studio を簡単に使用するのに役立つキーボード・ショートカットを示します。

機能	ショートカット・キー
保存	[Ctrl]+[S]
元に戻す	[Ctrl]+[Z]
検索	[Ctrl]+[F]
次を検索	[F3]
適用	[Alt]+[A]
リフレッシュ	[Alt]+[R]
カット	[Ctrl]+[X]
コピー	[Ctrl]+[C]
ペースト	[Ctrl]+[V]
予測	[Ctrl]+[P]
新規データ・レンズ	[Ctrl]+[N]
データ・レンズのオープン	[Ctrl]+[O]

## タブおよびサブタブ

タブは、類似の情報をサブタブ、ペインおよびテキスト入力ボックスを含む、読みやすくアクセスしやすい領域にグループ化します。タブは、ツールバーのすぐ下のワークスペースに表示され、任意の順序で起動できます。すべてのタブが常に使用可能なわけではありません。たとえば、「Translate」タブとサブタブは、翻訳ターゲットがアクティブ化されるまでは表示されません。

サブタブは、タブと同様に機能し、各タブに関連する特定の機能またはユーティリティを提供するため、タブごとに異なります。

Knowledge Studio に含まれるタブと関連するサブタブは、次のとおりです。

タブ	関連するサブタブ
Phrases (第2章「データ内のフレーズ」)	Define Phrases
	Define Items
	View Hierarchy
	Regression Test
Standardize (第3章「データの標準化」)	Standardize Terms
	Standardize Phrases
	Standardize Lines
	Unit Conversion
	Test Global Standardization
	Regression Test
Standardize Items (第4章「品目定義の標準化」)	Standardize Attributes
	Order Attributes
	Match Weights
	Test Attributes
	Test Item Standardization
	Regression Test

タブ	関連するサブタブ
Classify (第5章「データの分類」)	Classify from Data Classify from Item Definitions Classify from Rules Test Classification Regression Test
Translate (第6章「データの翻訳」)	New Phrases and Known Phrases New and Known Variable Term Phrases Reorder Test Translated Attributes Test Item Translation Test Global Translation Regression Test

## タスク・ペイン

対話型タスク・ペインを使用すると、ペインのタイプに固有の処理の実行が可能であり、これらの処理についてこのリファレンス全体で説明します。通常、Knowledge Studio に含まれるタスク・ペインは、次のとおりです。

### 階層構造

データがツリー形式の構造で表され、ノードがどのように関連しているかが示されます。すべての場合ではありませんが、ノードを他のペインにドラッグ・アンド・ドロップできます。親ノードを展開すると、関連するすべての子ノードを表示できます。

### フォーム

ナレッジを作成するためにデータをフィールドに入力し、オプションを選択します。

### グラフィック・ルール

データがグラフィック・アイコンで表され、このアイコンをドラッグ・アンド・ドロップするとデータを変更できます。

### 表形式

データが Microsoft Excel スプレッドシートと同様の表形式のフォームに表示されます。

### ウィザード

データがステップバイステップ方式の問合せで収集されます。

### 矢印

ペイン間の小さい上 / 下矢印を使用すると、ペインのサイズを変更できます。さらに、矢印をクリックするとそのペインが非アクティブになり、いずれかのペインを完全に展開してより詳細なデータを表示できます。非アクティブのペインを再表示するには、逆向きの矢印をクリックするとペインが再表示されます。

### コンテキスト依存メニュー

様々なコンテキスト依存 (ショートカット) メニューがあり、これらは、タスク・ペイン内のデータを右クリックすると Knowledge Studio のペインに表示されます。これらのメニューの内容についてはこのリファレンス全体で説明しますが、次の標準オプションが含まれる場合があります。

### フィルタおよびフィルタ解除

これらのオプションは、前述の説明のとおりデータのフィルタ処理とフィルタ解除の実行します。

**アイコン・ヘルプ**

表示される可能性のあるコンテキスト依存の各アイコンを説明します。

**ノードの展開**

選択したノードのすべてのサブノード（フレーズまたは用語）を階層形式で展開します。

**すべて展開**

選択したノードのすべてのサブノード（フレーズまたは用語）を階層形式で展開します。

**検索および再検索**

前述の説明のとおりテキスト・データを検索します。

**カテゴリ表示を削除**

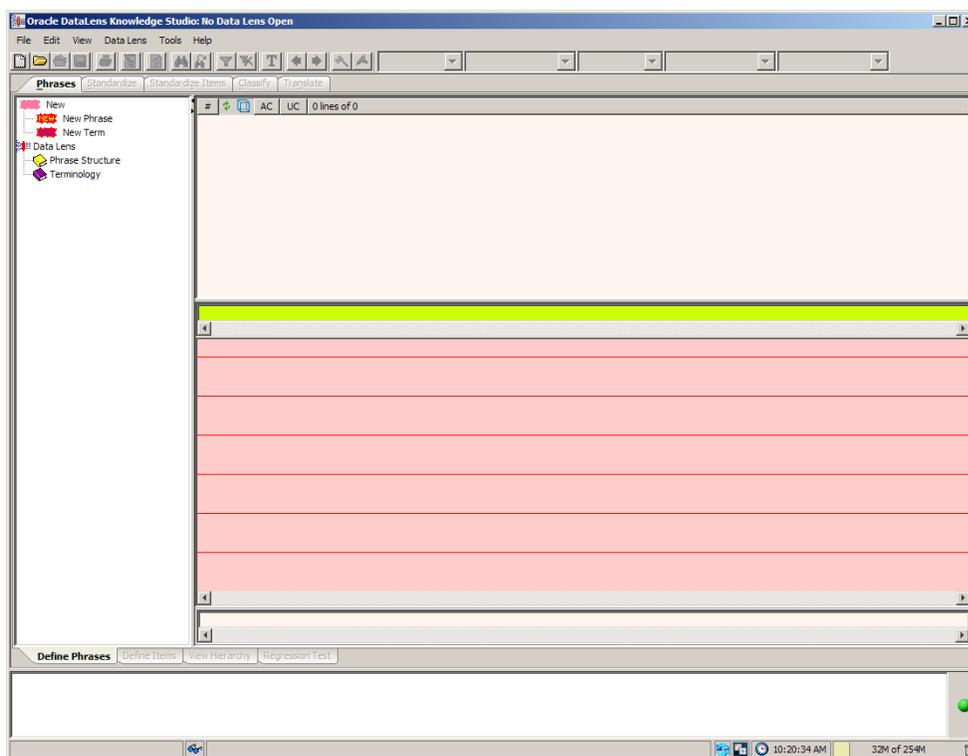
選択したカテゴリをペインの表示から削除します。カテゴリは、現在のデータ・レンズ編集セッションに対してのみ削除され、データ・レンズをクローズするとリセットされます。

**インターネットの検索**

選択したテキストをインターネットで検索します。これは、メニュー選択名の一部として表示されます。

## Knowledge Studio の起動

初めて Knowledge Studio を起動した場合、クライアント・ワークスペースは次の図のように空白で表示され、それ以外の場合は、前回のジョブ実行の結果が表示されます。



Knowledge Studio クライアント・ワークスペース下部にあるステータス・フィールドには、ロードしたデータ・レンズに関する情報が白いフィールドに表示され、日時およびメモリー使用がグレーのフィールドに表示されます。ステータス・フィールドは、最初の Knowledge Studio プロジェクトを作成するまでは空白であり、作成した時点でプロジェクトのステータスが表示されます。詳細は、1-4 ページの「[クライアント・ワークスペースの理解](#)」を参照してください。

## データ・レンズの作成またはオープン

Knowledge Studio を起動すると、オープンする既存のデータ・レンズを選択するためのプロンプトが表示されます。



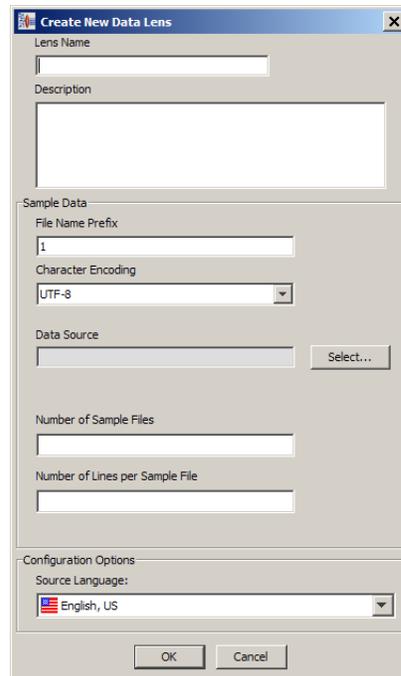
1. 新しいデータ・レンズを開始しているため、「**Cancel**」をクリックしてこのダイアログをクローズします。「**File**」メニューから、「**New Data Lens**」をクリックして新しいデータ・レンズを作成します。
2. このデータ・レンズに対する一意の名前を入力します。

---

**注意：** 空白を入力するとアンダースコアになります。

---

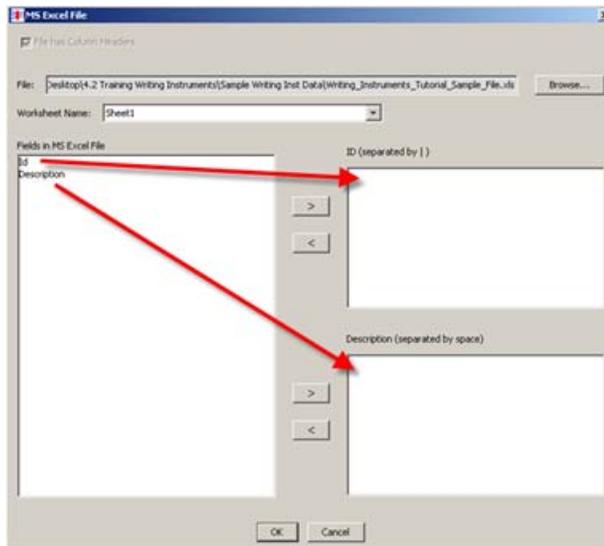
3. 概要を入力し、リストから「**Character Encoding**」を選択します。



4. 「**Data Source**」フィールドの横にある「**Select**」ボタンをクリックして、データが含まれているファイルを選択します。「**Data Source**」ダイアログが表示されます。



5. 「MS Excel file」オプションを選択し、「Specify」ボタンをクリックします。
6. 「Browse」をクリックし、データ・ファイルを検索してそのファイルを選択します。  
Excel ファイルのフィールド名が表示されます。
7. 「ID」フィールドを選択し、右矢印をクリックして「ID」リストを移入し、次に「Description」フィールドを選択して「Description」リストを移入します。「ID」は Excel スプレッドシートの部品番号フィールドに対応し、「Description」は、品目名やいくつかの属性などの部品の摘要です。



8. 「OK」をクリックします。「Data Source」ダイアログが表示され、指定したソース・ファイルとデータの行数が示されます。
9. 「OK」をクリックします。  
「New Data Lens」ダイアログに戻ります。
10. 「OK」をクリックします。Knowledge Studio によって、一連のサンプル・ファイルを含め、新しいデータ・レンズが作成されます。これらのサンプル・ファイルは、Excel スプレッドシートでのデータの XML 表現です。

新規プロジェクトは、次の場所に格納されます。

C:\Documents and Settings\Username\AppData\DataLens\data\project

または

C:\Users\Username\AppData\DataLens\export

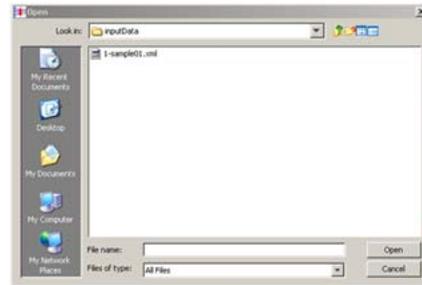
コンテンツおよびサンプル・ファイルは、次の場所に格納されます。

...\data lens name\inputData

サンプル・ファイルのファイル拡張子は、.xml です。



11. サンプル・データ・ファイルを選択するプロンプトが表示された場合は、「**Browse**」をクリックします。



12. サンプル・データ・ファイルを選択し、「**Open**」をクリックします。  
作成したデータ・レンズがオープンし、使用できる状態になります。



---

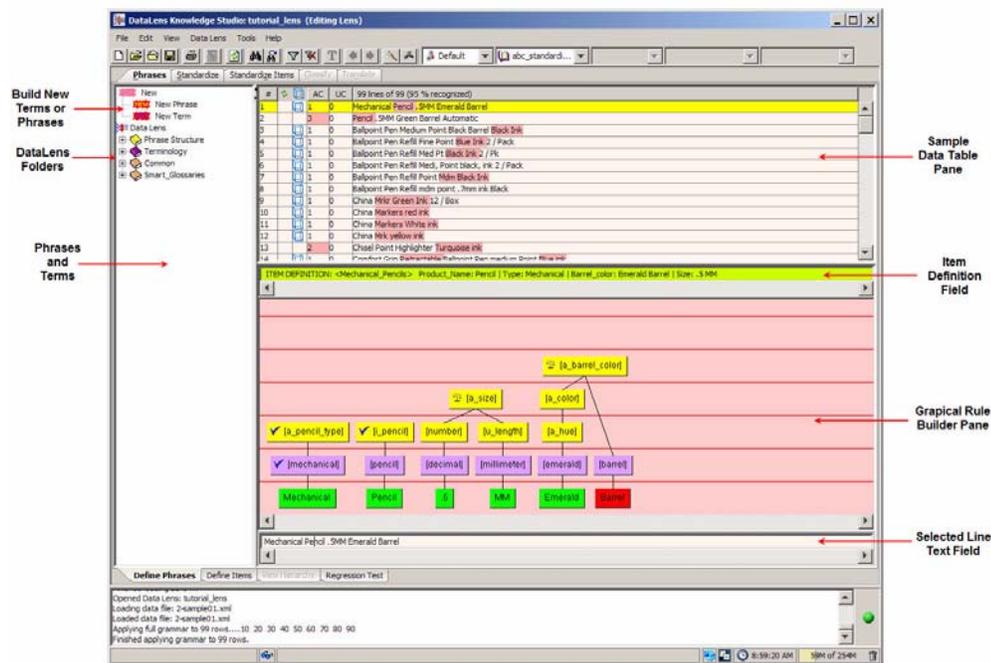
---

## データ内のフレーズ

この章では、用語ルールおよびフレーズ構造の作成方法について説明します。

## 「Phrases」 タブ

「Phrases」 タブおよび関連するサブタブを使用して、データ・レンズ内のサンプル・データを記述するフレーズ構造を作成します。「Define Phrases」サブタブがデフォルトで、データ・レンズの作成を開始する際に最初に使用するサブタブです。



## 「Define Phrases」 サブタブ

Knowledge Studio では、サンプル・データが自動的に区分されます。サンプル・データの選択された行は黄色に色付けされ、選択された行内で認識されないテキストは金色、あいまいなテキストはピンク色に色付けされます。

## 「Phrases and Terms」 ペイン

「Phrases and Terms」 ペインには、次に示すように、データ・レンズにフレーズ構造を作成するのに必要なすべてのコンポーネントが含まれています。

### New Phrase、New Term

この2つのアイコンを「Graphical Rule Builder」ペインにドラッグして、それぞれ新規のフレーズまたは用語ノードを作成できます。

### 「Data Lens」 フォルダ

このフォルダはデータ・レンズ全体の親フォルダで、データ・レンズに関連付けられたすべての構造が含まれています。これらのフォルダに含まれているノードを「Graphical Rule Builder」ペインにドラッグして、フレーズ構造を作成できます。

### 「Phrase Structure」 フォルダ

使用するフレーズ・ルールが格納されます。

### 「Terminology」 フォルダ

使用する用語ルールが格納されます。

### 「Common」 フォルダ

共通して使用するフレーズ構造および用語ノードが格納されます。

### 「Silver\_Creek\_Foundations」フォルダ

以前のバージョンの Enterprise DQ for Product で提供されたサンプル構造が格納されます。これは古いバージョンの「Smart Glossaries」フォルダで、インポート可能なレンズでこのドメイン名が使用されている場合に備えて存在します。基盤データ・レンズが存在しない場合、このフォルダは存在しません。

### 「Smart\_Glossaries」フォルダ

データ・レンズにインポートされたすべてのスマート用語集が格納されます。スマート用語集がインポートされていない場合、このフォルダは存在しません。

次のコンテキスト依存メニューは、「Phrases and Terms」ペインおよび「Graphical Rule Builder」ペインの両方で機能します。

#### Referenced By...

選択したルールを参照する品目定義、ルールおよび分類を含めて、ルール（フレーズまたは用語）に関する情報が表示されません。

#### Jump to Standardization

「Standardize」タブの「Standardize Phrases」または「Standardize Terms」サブタブがアクティブ化されて選択内容が表示され、品目定義を編集できます。

#### Review Productions

選択したルールまたは用語に関連付けられた生成結果をレビュー、移動または削除できるダイアログが表示されます。詳細は、2-10 ページの「[フレーズの生成結果の変更](#)」を参照してください。

Referenced By...	r
Filter Data	d
Jump to Standardization	j
Review Productions...	p
Rename Rule	a
Delete Rule	l
Merge Rule...	m
Edit Rule	e
Attributes	▶
Find...	f
Find Next (F3)	n
Icon Help	▶

#### Filter Data on rule name

選択したルールに基づいてサンプル・データをフィルタ処理し、「Sample Data Table」ペインを更新します。「Unfilter」ボタンを使用すると、データがフィルタ処理される前の状態に戻ります。

#### Merge Rule...

選択したルールを別のルールにマージできるダイアログが表示されます。詳細は、2-9 ページの「[2つのフレーズ・ルールのマージ](#)」を参照してください。

#### Rename Rule

選択したフレーズまたは用語ルールの名前を変更できるダイアログが表示されます。詳細は、2-9 ページの「[フレーズ・ルールの名前変更](#)」を参照してください。

#### Delete Rule

確認メッセージが表示され、ルールとともに関連するルールや生成結果もすべて削除されることを含めて、選択したルールを削除した場合の結果をレビューするように求められます。

#### Edit Rule

選択したルールを編集できるダイアログが表示されます。詳細は、2-10 ページの「[フレーズ・ルールの編集](#)」を参照してください。

#### Create New Parent

新規の親用語を作成します。詳細は、2-13 ページの「[親ノードの作成、挿入または切断](#)」を参照してください。

#### Disconnect from Parent

下位レベルのノードと親ノードの間の接続を削除します。

#### Insert New Parent

選択したノードに対して新規の親用語を挿入します。詳細は、2-13 ページの「[親ノードの作成、挿入または切断](#)」を参照してください。

#### Attributes

次のコンテキスト依存メニューの説明を参照してください。

次の「Attribute Rule」コンテキスト依存メニューはフレーズ構造や用語ノードによって多少異なりますが、メニュー・オプションは同様に機能します。各種の属性オプションを設定すると、データ・レンズ処理が次のように変わります。

#### Do Not Translate

ルールは「Translation」タブに表示されず、翻訳の選択対象になりません。詳細は、6-1 ページの「データの翻訳」を参照してください。

#### Format to Locale

ターゲット言語の要件に基づいて、選択内容が自動的に書式設定されます。

#### Translation Variable

ルールは1回のみ翻訳され、その後は翻訳時に再利用される変数として使用されることを示します。

#### Prohibit Rename

ルールの名前を変更できないようにします。

#### Prohibit Anchoring

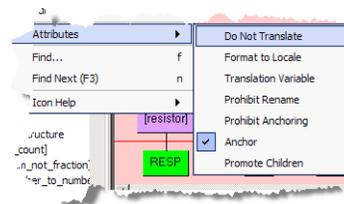
ルールはすべてのフレーズに固定できません。

#### Anchor

ルールをフレーズに固定します。これがデフォルトです。

#### Promote Children

親属性と同じ値を使用できるように、ルールのすべての子を昇格します。



## 「Sample Data Table」 ペイン

「Sample Data Table」ペインには、選択したサンプル・データ・ファイルのサンプル・データの各行が表示されます。表の各列は、左から右に次の内容を示しています。

#### 行番号 (#)

データ行に割り当てられている一意の番号。

#### リフレッシュ

データ行をリフレッシュする必要がある場合は、この列にリフレッシュ・アイコンが表示され、「Phrases」タブもリフレッシュ・アイコンでマークされます。リフレッシュによって、ロードされているサンプル・データ・ファイルのすべての行に変更内容が適用されます。特定のデータ行にまだ適用されていない文法のルールがある場合は、この列にリフレッシュ・アイコンが表示されます。一度に1行のみをリフレッシュするにはその行をクリックし、すべてのサンプル・データをリフレッシュするには「Refresh」ボタンをクリックします。

#### あいまい表現件数 (AC)

データ行のあいまい表現件数を示し、あいまい表現を解決できるように対応するテキストはピンク色で色付けされます。詳細は、2-13 ページの「フレーズのあいまい表現の解決」を参照してください。

#### 未解析件数 (UC)

行内で未解析の用語を示し、対応するテキストは金色で色付けされます。これによって、作成する必要がある用語およびフレーズのルールを簡単に識別できるため、データを完全に解析できます。

#### 認識された行の数 / 割合

列ヘッダーに、データ・レンズによって認識された列の数とデータの割合が表示されます。これは、データがリフレッシュされたり、ルールが作成または編集されると更新されます。

データが含まれている各列は、そのタイトルをクリックすることで表を昇順と降順の両方でソートできます。列ヘッダーを1回クリックすると、表がソートされ、選択した列の品目別に

昇順のアルファベット順で表示されます。同じ列ヘッダーを2回クリックすると、表が再度ソートされ、降順のアルファベット順で表示されます。

次のコンテキスト依存メニューは「**Sample Data Table**」ペインで機能し、各オプションを使用して次の処理を実行できます。

#### Mark Lines for Delete

選択したすべての行の「#」列が赤色に変わりますが、行はまだ削除されません。

#### Un-Delete Lines

削除対象としてマークされた行を通常の状態に戻します。

Mark Lines for Delete	m
Un-Delete Lines	u
Append	
Mark Lines for Delete	m
Un-Delete Lines	u
Apply Deletes	d
Append Sample Data	s

#### Apply Deletes

削除対象としてマークされたすべてのデータ行を完全に削除します。このオプションは、削除対象としてマークされた行がある場合のみアクティブになります。削除された行は取得できなくなるため、このオプションは慎重に使用してください。

#### Append Sample Data

追加のサンプル・データをデータ入力行から現在のデータ・ソース・ファイルに追加します。

## 「Item Definition」フィールド

このフィールドには、「**Sample Data Table**」ペインで選択した行の品目定義が表示されます。このフィールドはレビュー専用です。

## 「Graphical Rule Builder」ペイン

このペインを使用して、データ行のテキスト・コンポーネントを識別するフレーズおよび用語構造を視覚的に表現できます。選択したデータ行に対してルールを作成または変更するには、フレーズまたは用語ルールを「**Phrases and Terms**」ペインから「**Graphical Rule Builder**」ペインにドラッグ・アンド・ドロップします。

次のコンテキスト依存メニューは、「**Graphical Rule Builder**」ペインで機能します。

#### Predict Terms

予測しきい値を満たした信頼度に基づいて、各サンプル行の入力データに適用可能なすべてのルールが表示され、ルールを選択できます。予測が表示されない場合は、その理由を示すメッセージが表示されます。「**Predict Terms**」は、サンプル行に品目定義が関連付けられている場合に、品目定義のコンテキストでのみ機能します。適切なルールを選択するか、または予測を拒否できます。予測の拒否は、現在のデータ・レンズ編集セッションに対してのみ適用可能で、データ・レンズをクローズするとリセットされます。このオプションは「**Edit**」メニューにもあります。詳細は、2-11 ページの「[ルール予測の使用](#)」を参照してください。

Predict Terms	t
Predict Best	b
Predict Required Attributes	r
Predict Missing Attributes	m
Predict All	a
Undo Predict All	x
Layout Nodes	l

#### Predict Best

予測しきい値を満たした信頼度に基づいて、選択したレコードの未解析テキストに対して最も近似一致するフレーズ・ルールが自動的に選択されて表示され、ユーザーがルールを選択することもできます。予測が表示されない場合は、その理由を示すメッセージが表示されます。「**Predict Best**」は、サンプル行に品目定義が関連付けられている場合に、品目定義のコンテキストでのみ機能します。詳細は、2-11 ページの「[ルール予測の使用](#)」を参照してください。

#### Predict Required Attributes

選択したレコードの未解析テキストに対して、品目定義の必須属性と一致するフレーズ・ルールのみが表示されます。詳細は、2-11 ページの「[ルール予測の使用](#)」を参照してください。

**Predict Missing Attributes**

品目定義に指定した属性順序に基づいて、選択したレコードの未解析テキストに対して一致するフレーズ・ルールのみが表示されます。詳細は、2-11 ページの「[ルール予測の使用](#)」を参照してください。

**Predict All**

全レコードの未解析データをすべて検証し、すべての予測を受け入れる場合に、このオプションを選択できます。詳細は、2-11 ページの「[ルール予測の使用](#)」を参照してください。

**Undo Predict All**

「Predict All」 オプションを使用して変更された内容を元に戻します。このオプションは「Predict All」が選択された後のみ有効になります。詳細は、2-11 ページの「[ルール予測の使用](#)」を参照してください。

**Layout Nodes**

「Graphical Rule Builder」 ペイン内のノードを再編成して再表示します。これは、ルールに対して複数の変更を加えた場合に非常に便利です。

**New Term**

新規用語を追加します。詳細は、2-7 ページの「[不明品目に対する用語ルールの作成](#)」を参照してください。

New Term	n
Change to Regex	r

**Change to Regex**

選択したノード（用語ルールの生成結果）を正規表現に変更します。これは、数値や記号の場合、または空白を含まないテキストが文字列内にある場合に最も便利です。詳細は、[付録 B 「正規表現」](#)を参照してください。

**「Selected Line」 フィールド**

「Selected Line」 フィールドを使用して、選択したデータ行を編集し、その変更内容をサンプル・データに追加できます。このフィールド内のデータは、「Graphical Rule Builder」 ペインに表示されているデータと同じです。このフィールドでは、サンプル・データの選択されている行を区分して、さらに詳細なフレーズ・ルールを作成できます。

次のコンテキスト依存メニューは、「Selected Line」 フィールドで機能します。

**Add to Sample Data**

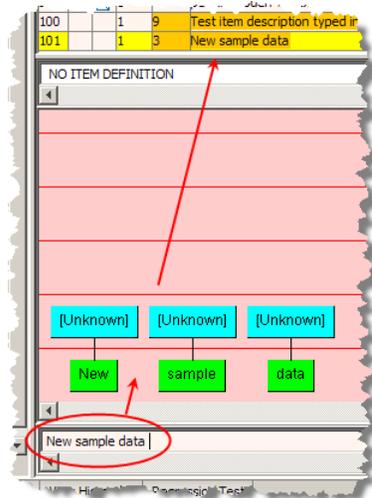
Add to Sample Data

「Selected Line」 フィールドで行った編集内容をすべてサンプル・データに追加します。

「Selected Line」 フィールドを使用して、トライアル・コンテンツ、つまりサンプル・データ・ファイルに含まれていない内容を入力できます。たとえば、このフィールドに「New sample data」と入力し、「Add to Sample Data」 オプションを使用します。これによって、次の図に示すように、入力したテキストが新規用語として追加されます。

IDを入力するか、またはフィールドを空白のままにして、「OK」をクリックします。

サンプル・データ・ファイルの最後の行に新しい行が表示されます。



## ツールの使用

次の各項では、ツールの使用方法について説明します。

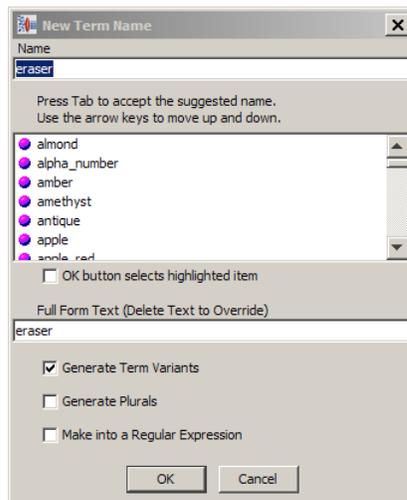
### 不明品目に対する用語ルールの作成

不明品目に対して用語ルールを作成するには、次の3つの方法があります。

- 緑色のテキスト・ノードを右クリックして表示されるコンテキスト依存メニューから、「**New Term**」をクリックします。
- 青色の「**[Unknown]**」ノードを右クリックし、「**New Term**」をクリックします。用語名の入力を求められます。「mechanical」というテキストに対する用語ルールを作成します。
- 緑色のテキスト・ノードをダブルクリックして、同じ名前の用語ルールを作成します。

ダブルクリックする方法を使用できるのは、テキストと完全一致する名前を用語ノードの名前として使用する場合のみです。これによって、この用語ルールの完全形式のデフォルト値も提供されます。

どの方法を使用しても、「**New Term Name**」ダイアログが表示されます。



1. 新規用語ルールのわかりやすい名前を入力します。
2. 適切なチェック・ボックスを選択します。

#### Generate Term Variants

用語ルールの作成時に、ルールの用語バリエーションが自動的に生成されます。たとえば、`rsstr` は `RESISTOR` のバリエーションです。

#### Generate Plurals

完全形式の用語のバリエーションを認識するために、用語ルールの作成時に、ルールの複数形が自動的に生成されます。

#### Make into a Regular Expression

用語ルールの生成結果を、完全形式の用語の正規表現に変換します。通常、これは数値または記号に対してのみ有効です。

3. 「OK」をクリックして新規用語を追加します。新規の用語ノードが「Phrases and Terms」ペインに表示されます。

用語ルールを作成する際のベスト・プラクティスは、完全形式を入力することです。これによって、予測機能が強化されます。予測の詳細は、2-11 ページの「[ルール予測の使用](#)」を参照してください。

## 用語ノードに対するフレーズ構造の作成

この項では、用語ルールが含まれる上位レベルのフレーズ・ルールの作成方法について説明します。適格なフレーズ構造には、1つ以上の用語ノードと1つ以上のフレーズ・ノードが含まれます。

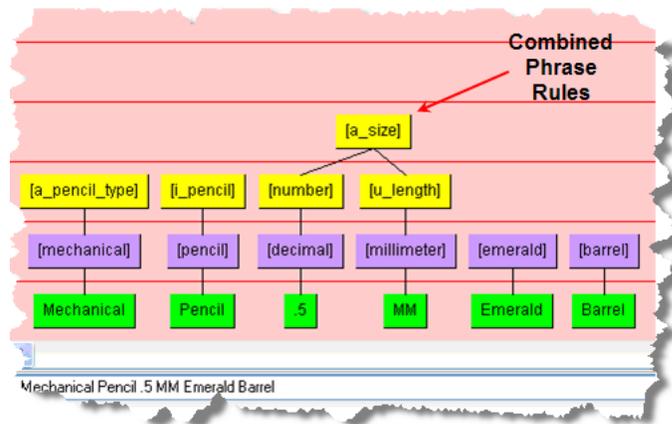
新規のフレーズ構造を作成するには、「Phrases and Terms」ペインの上部にある「New Phrase」アイコンを既存のフレーズ・ノードまたは用語ノードにドラッグするか、用語ノードを右クリックして「Create New Parent」をクリックするか、または用語ノードをダブルクリックします。前述の説明に従って、新規ルールを作成します。

## 2つのフレーズ・ルールを1つのフレーズ・ルールに結合

2つのノードを結合すると、1つのセマンティック概念として結合して、データ・レンズのフレーズ・ナレッジを統合できます。

次の手順を使用して、テキスト・ノードを結合して1つの生成結果を形成します。

1. 新規フレーズを用語ノードの親として作成します。  
ここでは汎用的な測定のフレーズを作成するため、他の解析ルールでも使用できる名前を割り当てることをお勧めします。
2. 用語 `[millimeter]` を選択して右クリックし、「Create New Parent」をクリックします。
3. `[number]` を選択し、フレーズ `[a_size]` を作成します。
4. `[a_size]` フレーズ・ノードを `[u_length]` フレーズ・ノードに接続するには、下位レベルのノード `[u_length]` をクリックします（赤色に変わります）。
5. `[a_size]` の中央をクリックし、フレーズの黄色の丸いアイコンをクリックして `[u_length]` にドラッグします。

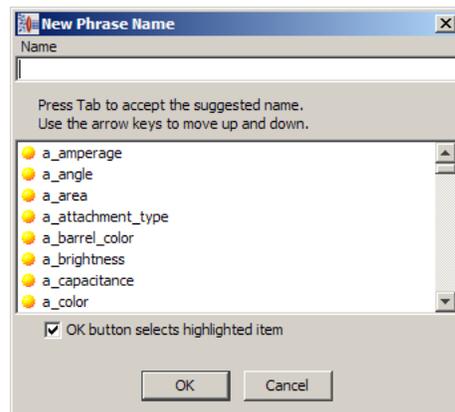


これで、フレーズ構造を定義する線が2つのノードを結合して、1つの結合されたフレーズ構造が形成されます。

## 2つのフレーズ・ルールのマージ

ルールをマージすると、2つのルールの生成結果を1つのルールに結合する場合に便利です。次の手順を使用して、ルールをマージします。

1. マージする用語またはフレーズを右クリックし、「Merge Rule...」をクリックします。



2. マージする用語またはフレーズを検索し、[Tab] キーを押してその用語またはフレーズを選択します。

スクロールバーを使用してマージする用語またはフレーズを検索するか、または「Name」フィールドに名前を入力して検索します。

3. 「OK」をクリックしてルールをマージします。

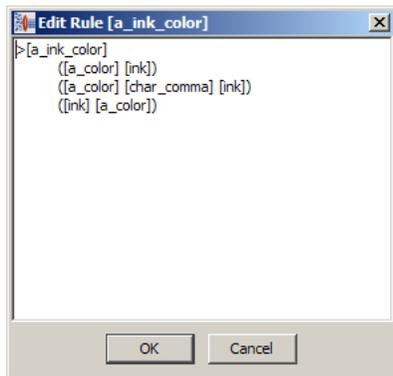
ルールのマージ操作の結果を表示するには、用語またはフレーズを右クリックし、「Review Productions」をクリックします。

## フレーズ・ルールの名前変更

フレーズ・ルールの名前は、「Graphical Rule」ペインでフレーズを右クリックして「RenameRule」を選択することで簡単に変更できます。「New Term Name」ダイアログを使用して、このフレーズ・ルールの名前を変更します。詳細は、2-7 ページの「不明品目に対する用語ルールの作成」を参照してください。

## フレーズ・ルール編集

編集するルールを右クリックし、「Edit Rule」を選択します。



「Edit Rule」ダイアログでは、テキスト・エディタと同様の操作で、選択したルールを直接編集できます。このダイアログを使用して、既存の生成結果の構文が正しいことを検証し、必要に応じて編集して「OK」をクリックします。

## 用語のバリエーションの追加

最も強力なフレーズおよび用語ルールを作成するには、サンプル・データが迅速かつ標準的な方法で解析されるように、用語ノードに関連付けることが可能なバリエーションを追加します。

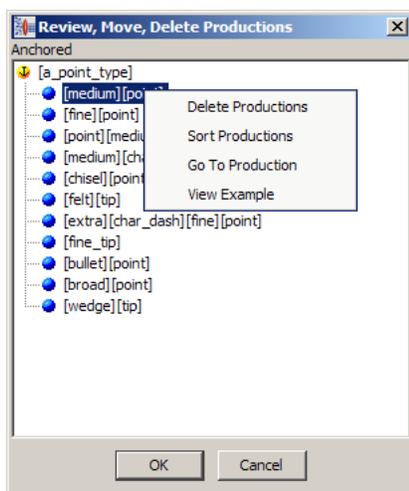
たとえば、用語 **[medium]** は、用語「med」が出現した場合にこれを用語バリエーションとして追加するように定義できます。

用語のバリエーションを生成するには、前述の説明に従って用語を編集し、「Generate Term Variants」を選択します。

## フレーズの生成結果の変更

ルールに対して作成されたすべてのフレーズの生成結果が表示されます。

ルールを選択して右クリックし、「Review Productions」を選択します。次に、**[a\_point\_type]** がデータ内で認識される3つの方法を示します。



生成結果を削除するには、生成結果を選択して右クリックし、「Delete Productions」をクリックします。その他のコンテキスト依存オプションを使用すると、次の処理を実行できます。

### Sort Productions

リストされている生成結果を昇順のアルファベット順にソートします。

### Go to Production

「Standardize」タブの「Standardize Phrases」サブタブで、選択した生成結果がアクティブ化されます。

### View Example

編集可能フィールドに選択した生成結果のテキスト例が表示され、編集できます。

## ルール予測の使用

フレーズおよび用語を既存のルールに追加する便利な方法は、Knowledge Studio の予測機能を使用することです。予測オプションを使用して、選択したレコードまたは全レコードの未解析ソース・テキストを検証し、信頼度に基づいてテキストが最も近似一致するルールを判別します。

予測オプションはデータ・レンズ内の品目定義に基づいてアクティブ化されます。次に詳細を説明します。

### Predict Terms

予測しきい値を満たした信頼度に基づいて、各サンプル行の入力データに適用可能なすべてのルールが表示され、ルールを選択できます。予測が表示されない場合は、その事由を示すメッセージが表示されます。「Predict Terms」は、サンプル行に品目定義が関連付けられている場合に、品目定義のコンテキストでのみ機能します。適切なルールを選択するか、または予測を拒否できます。予測の拒否は、現在のデータ・レンズ編集セッションに対してのみ適用可能で、データ・レンズをクローズするとリセットされます。

**ヒント：** このオプションは「Edit」メニューにもあります。

### Predict Best

予測しきい値を満たした信頼度に基づいて、選択したレコードの未解析テキストに対して最も近似一致するフレーズ・ルールが自動的に選択されて表示され、ユーザーがルールを選択することもできます。予測が表示されない場合は、その事由を示すメッセージが表示されます。「Predict Best」は、サンプル行に品目定義が関連付けられている場合に、品目定義のコンテキストでのみ機能します。

### Predict Required Attributes

選択したレコードの未解析テキストに対して、品目定義の必須属性と一致するフレーズ・ルールのみが表示されます。変更内容は、表示されている品目定義の必須属性にのみ反映されます。

### Predict Missing Attributes

品目定義に指定した属性の順序付けルールに基づいて、選択したレコードの未解析テキストに対して一致するフレーズ・ルールのみが表示されます。品目定義に属性の順序付けルールが指定されていない場合は、すべての属性が検証されます。

### Predict All

全レコードの未解析データがすべて検証され、すべての予測が自動的に受け入れられて、データ・レンズに即時に反映されます。結果は「Status Field」に表示されます。

---

**注意：** この処理では確認メッセージが表示されないため、データ・レンズへの反映を元に戻すには「Undo Predict All」オプションを使用する必要があります。

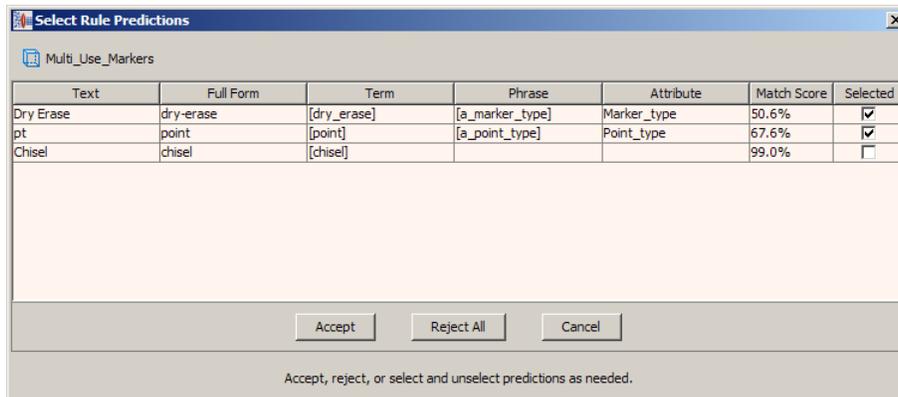
---

### Undo Predict All

「Predict All」オプションによって反映された変更内容をすべて元に戻す場合に、このオプションを選択します。「Predict All」と同様に、「Undo Predict All」を選択すると、確認メッセージは表示されずに、変更内容が即時に元に戻され、予測のキャッシュは消去されます。結果は「Status Field」に表示されます。

**注意：** 予測結果は、「DataLens Options」の「Prediction」タブで設定した予測しきい値を満たしているかどうかに基づきます。0%に設定すると使用可能な予測の完全セットが作成されますが、このように低く設定すると誤った予測も検出される可能性があります。また、設定値を非常に低く設定して「Predict All」を使用すると、不正な関連付けを誤って追加する可能性があります。「Predict All」の結果は慎重にレビューしてください。予測オプションの構成の詳細は、8-2 ページの「データ・レンズ・オプションの設定」を参照してください。

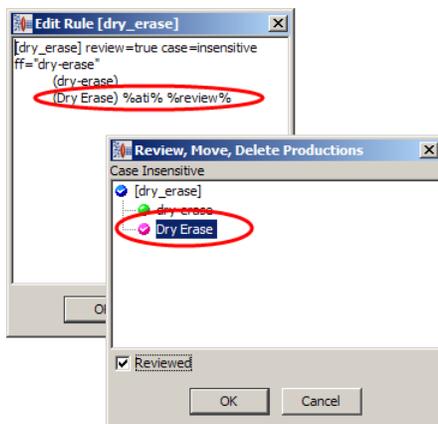
これらのオプションは「Define Phrases」サブタブでのみ使用可能で、「Sample Data」ペインでレコードを選択し、「Graphical Rule Builder」ペインを右クリックすると表示されます。



「Predict All」および「Undo Predict All」を除くすべてのオプションで、予測が検出されると「Select Rule Predictions」ダイアログが表示されます。表示されるデータの列は、元の未解析テキスト、テキストと一致している可能性がある完全形式の用語バリエーション、一致した用語、一致したフレーズ、一致した属性、および全体の信頼度の一致スコア（率）です。

可能な場合は、一致スコアが最も高い予測が選択されるため、そのまま受け入れることができます。ルール予測を選択して、受け入れるか、または拒否できます。予測の拒否は、現在のデータ・レンズ編集セッションに対してのみ適用可能で、データ・レンズをクローズするとリセットされます。

予測を受け入れると、用語ルールが更新され、その予測は「%ati% %review%」が追加されて追加されます。次の例に示すように、用語の生成結果をレビューするとき、その予測はピンク色で表示されます。



「Reviewed」チェック・ボックスを選択してルール予測をレビューしたことを示すと、ルールに付加された %review% が削除され、生成結果がピンク色の丸いアイコンから緑色の丸いアイコンに変わります。

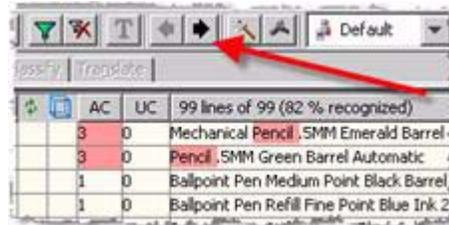
## フレーズのあいまい表現の解決

データ行内のテキスト（1つ以上のテキスト・ノード）の文字列を認識するフレーズ構造を複数作成することが可能です。フレーズ構造間で競合がある場合は、あいまい表現が発生します。

あいまい表現が存在する場合、解析結果の順序と内容は保証されません。

あいまい表現件数が1を超える場合は、データ行を認識するルールが複数あることを意味します。あいまい表現件数が1を超えると、この列はピンク色で表示され、用語および用語ルールとフレーズ・ルールは白色になります。

「Next」 / 「Previous」 ツールバー・ボタン（左矢印と右矢印）がアクティブになります。これらを使用して、行の内容と一致するフレーズ構造を表示できます。



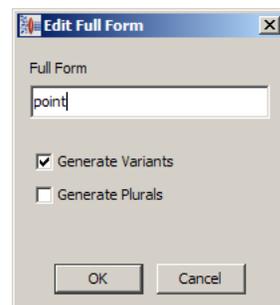
データ行の表示されていない部分にあいまい表現が発生している場合があるため、タスク・ペインの下部にある横方向のスクロールバーを使用して、データ行全体を表示する必要があります。

ルールをマージ、結合または削除して（あるいはこれらを組み合わせ）、発生したあいまい表現を解決すると、サンプル・データが正確になります。

あいまい表現を解決すると、「Sample Data Table」 ペインでは、あいまい表現件数の値が1に戻り、あいまいであった用語はピンク色で強調表示されなくなります。さらに、「Next」 / 「Previous」 ボタンが非アクティブになります。

## 用語の完全形式の編集

「Edit Full Form」 オプションを使用して、選択した用語の完全形式を変更したり、前述のように用語バリエーションや複数形を生成できます。



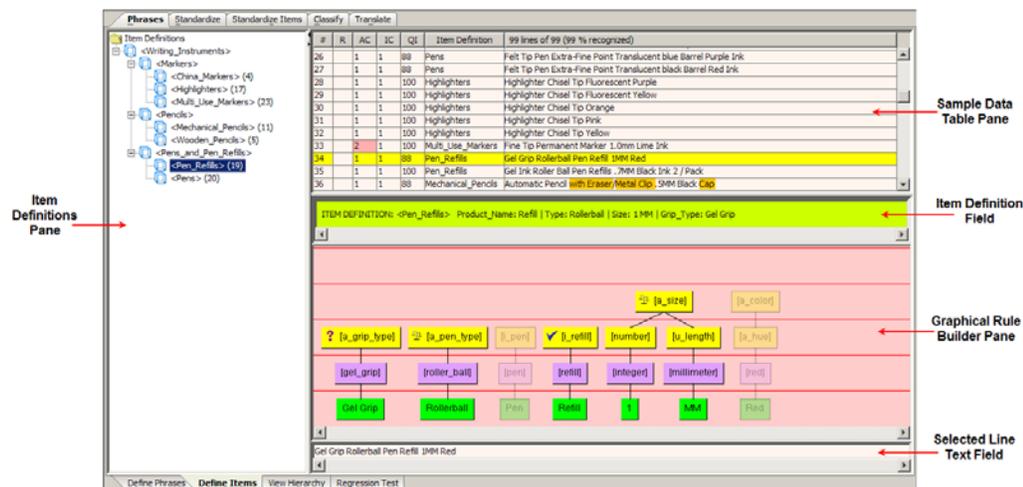
**親ノードの作成、挿入または切断** これらの機能を使用して、新規の親を追加したり、既存のノード間に新規の親ノードを挿入できます。ノードを選択して右クリックし、該当する機能を選択します。

新規の親を追加または挿入するときは、「New Phrase Name」 ダイアログが表示され、前述の説明に従って使用します。

親ノードからノードを切断すると、2つのノードを接続していた線が「Graphical Rule Builder」 ペインに表示されなくなります。ただし、2つのルール間の関係はデータ・レンズ内に残っています。親からの生成結果である子を削除する場合は、ルールを編集する必要があります。

## 「Define Items」 サブタブ

品目定義は、属性、属性の値、およびレンズ内の品目定義間の関係に基づいています。データ・レンズの品目定義では、属性に関連するオブジェクトを定義できます。品目定義は、階層スキーマまたは分類に明示的または暗黙的に基づく場合もあります。品目定義によって、定義された製品およびその属性に最上位レベルのドメインからアプローチできます。



## 「Item Definitions」 ペイン

「Item Definitions」 ペインには、データ・レンズに定義されたすべての品目が表示されます。フォルダ内は階層構造になっており、最上位レベル品目定義からすべての下位レベル品目定義まで製品分類が表示されます。

次のコンテキスト依存メニューは、このペインで使用できます。

### Expand Node

選択したノードの階層構造全体が表示されます。

### Create Top-Level Item Definition

最上位レベルの定義を作成します（この項で後述）。

### Create Sub-Item Definition

最上位レベルの定義を作成します（この項で後述）。

### Edit Item Definition

「Item Definition Attribute」 ペインがアクティブ化され、選択したノードを編集できます。詳細は、2-20 ページの「「Item Definition List」 ペイン」を参照してください。

### Edit Item Definition Alias or Description

別名または摘要を編集できます（この項で後述）。

### Copy Item Definition

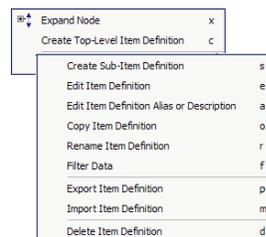
選択した品目定義を、同じレベルで、指定した名前を使用して新規の品目定義にコピーします。

### Rename Item Definition

選択した品目定義の名前を、指定した名前に変更します。

### Filter Data

選択内容に基づいて、「Sample Data Table」 ペイン内のサンプル・データをフィルタ処理します。



**Set Inactive for Production**

選択した品目定義を非アクティブに設定できます。詳細は、2-18 ページの「[アクティブまたは非アクティブな品目定義](#)」を参照してください。

**Activate this and all children for Production**

本番で使用できるように、選択した品目定義とそのすべての子品目定義をアクティブにします。詳細は、2-18 ページの「[アクティブまたは非アクティブな品目定義](#)」を参照してください。

**Set Inactive this and all children for Production**

選択した品目定義を非アクティブに設定できます。詳細は、2-18 ページの「[アクティブまたは非アクティブな品目定義](#)」を参照してください。

**Export Item Definition、Import Item Definition**

品目定義をデータ・レンズからエクスポートしたり、データ・レンズにインポートします。詳細は、9-4 ページの「[品目定義のエクスポートとインポート](#)」を参照してください。

**Delete Item Definition**

選択した品目定義の削除を確認する確認メッセージが表示されます。

**Delete Item Definition and Children**

選択した品目定義とそのすべての子品目定義の削除を確認する確認メッセージが表示されます。

**「Sample Data Table」ペイン**

列は、「[Define Phrases](#)」サブタブ (2-2 ページの「[「Define Phrases」サブタブ](#)」を参照) の「[Sample Data Table](#)」ペインの列に似ており、同じように動作します。

**品目定義件数 (IC)**

データ行で認識された品目定義の数を示します。

**品質指標 (QI)**

データ行の品質指標を示します。これは、品目定義に関連付けられた必須属性とスコアリング属性の数に基づく、属性カバレッジの測定値です。

QI は、品目定義に属性が存在することに対して 50 ポイント、必須属性とスコアリング属性の合計が占める割合に対して 50 ポイントとして、次のように計算されます。

$$50 + 50 * (\text{必須属性とスコアリング属性の合計数} / \text{属性合計数})$$

たとえば、属性合計数が 10 で、2 つの必須属性と 3 つのスコアリング属性がその中に含まれる場合、次の計算によって QI は 75 になります。

$$50 + 50 (5/10) = 75 \text{ QI}$$

データ行の QI が 100 の場合は、品目定義に必須属性とスコアリング属性の各属性値が含まれています。

**Item Definition**

データ行の品目定義の名前が表示されます。

**「Item Definition」フィールドおよび「Selected Line」フィールド**

これらのフィールドは、「[Define Phrases](#)」サブタブ (2-2 ページの「[「Define Phrases」サブタブ](#)」を参照) にあるフィールドと同じです。

## 「Graphical Rule Builder」 ペイン

「Graphical Rule Builder」 ペインは、「Define Phrases」 サブタブ (2-2 ページの「「Define Phrases」 サブタブ」を参照) にある同じペインと同様に機能します。ただし、この「Graphical Rule Builder」 ペインで使用できるコンテキスト依存メニューは、次のように異なっています。

### Referenced By

選択したルールを参照する品目定義、ルールおよび分類を含めて、ルール (フレーズまたは用語) に関する情報が表示されます。

Referenced By	r
---------------	---

### Match Attributes

品目定義の属性と一致する、データ行のフレーズ構造が表示されます (たとえば、{a\_color} フレーズは「Color」属性と一致します)。

Match Attributes	m
Reverse Ghosting	r

### Reverse Ghosting

トグルとして機能し、品目定義に関連付けられていない非アクティブ (不明瞭) な用語を表示します。これがデフォルトの動作です。このオプションを選択すると用語がアクティブとして表示され、再度選択すると、動作がデフォルトにリセットされます。

## ツールの使用

次の各項では、ツールの使用方法について説明します。

### 新規の最上位レベル品目定義の作成

次の手順を使用して、新規の品目定義を作成します。

1. 新規の品目定義を作成するには、左ペインにある「Item Definitions」フォルダを右クリックし、「Create Top-Level Item Definition」をクリックします。

2. 品目定義の名前を入力します。
3. (オプション) 別名 (ラベル) を入力します。
4. (オプション) 摘要を入力します。
5. 「OK」をクリックします。

新規の品目定義が、左側の「Item Definition」ペインの「Item Definition」フォルダに表示されます。

### サブ品目定義の作成

サブ品目定義の作成方法は前述の項で説明したプロセスと同じですが、最初に親品目定義を選択してから「Create Sub-Item Definition」を選択する必要があります。

## 別名

次の各項では、品目定義および属性の別名の使用について説明します。

**別名を使用した品目定義および属性のラベル付け** 別名は、品目定義および属性の名前のラベルとして便利です。特に、ダウンストリーム処理でラベルに正確に反映する場合に（品目定義や属性の名前で使用できない空白、大文字表記、数字および特殊文字を含む）、別名を使用します。

データ・レンズは他のプロセスに入力を提供するため、別名を使用すると、別名によるラベルとデータで使用されている標準ラベルを照合するのに役立ちます。

**別名の制限** 別名には、ほとんどすべての文字を使用できます。属性の場合は、品目定義ごとに同じ別名を複数回使用できます。さらに、品目定義で属性が使用されるすべての箇所（複数回使用される場合があります）、同じ別名を使用する必要があります。データ・レンズ全体で、属性に対する同じ別名を保守する必要があります。

別名が命名要件に従っていないと、エラー・メッセージが表示されるか、または、他にエラーがない場合は別名なしで品目定義または属性が作成される場合があります。このような場合は、品目定義または属性を作成した後に、別名を編集して許容可能な別名を追加できます。

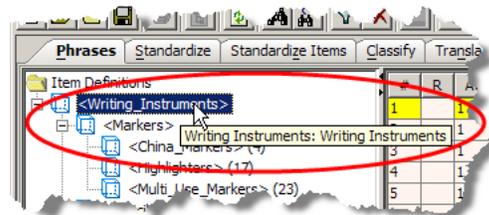
**別名の編集** 別名を編集するには、別名でラベル付けされた品目を右クリックし、「Edit Alias」（属性の場合）または「Edit Item Definition Alias or Description」（品目定義の場合）を選択して、表示されたダイアログでラベル（または他の情報）を変更します。品目定義の別名と摘要は階層リストからのみ編集できるのに対して、属性の別名は「Item Definition Edit」ペインの属性リストから編集できます。

別名は必須ではありません。別名を指定しない場合は、属性名が使用されます。

**別名と属性名が異なる場合の優先ルール** 別名が属性名と同じでない場合、出力では別名によって属性名が上書きされます。次の例では、属性 Att\_1\_Name に別名 Pen が指定されているため、スプレッドシートの H 列は次の図のように表示されます。

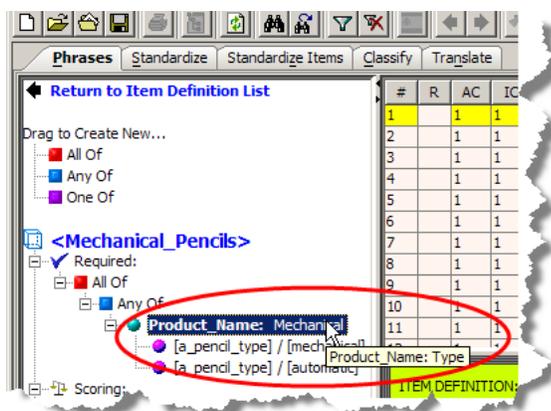
	H	I	J
1	Att_1_Name	Att_1_Value	Att_1_Text
2	Pen		
3	Pen		
4	Pen		
5	Pen		
6	Pen	GEL	GEL
7	Pen	INK	INK
8	Pen		
9	Pen	LIQUID	LIQUID
10	Pen	GEL	GEL
11	Pen	GEL	GEL

**別名情報の表示** 別名情報を作成または編集するときに、別名情報を表示できます。さらに、次の例に示すように、オブジェクトの上にマウスをあわせて、ツールヒントに別名情報を表示することもできます。



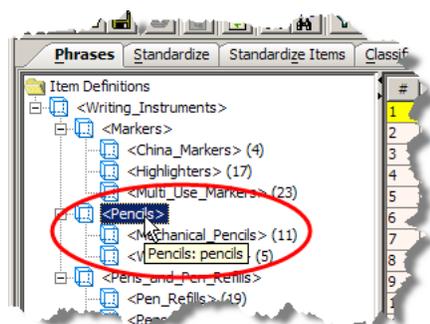
品目定義の別名の後に品目定義の摘要が表示されます。

属性の別名情報は、次の図のように表示されます。



属性に摘要も設定されている場合、その情報は別名情報の後に表示されます。属性に対して摘要のみが設定されている場合は、属性の上にマウスをあわせると摘要のみが表示されます。

別名は、次の図に示すように、「Item Definition List」ペインにマウスをあわせることによって表示されます。



## アクティブまたは非アクティブな品目定義

AutoBuild を使用してソース・データから品目定義を作成すると未完了になる場合があるため、本番使用のためにデータ・レンズにチェックインする前の確認時に品目定義が十分にレビューされないと、未完了の品目定義に対して大量の一致が検出される結果になります。これは、品目定義を作成して、フレーズおよび用語ルールを移入している最中にも発生する場合があります。

アクティブおよび非アクティブ・フラグに基づいて、データ・レンズ内の品目定義を出力データに含めるように設定できます。Knowledge Studio では、すべての品目定義を使用してデータを解析しますが、非アクティブ・フラグが設定されている場合、本番モードで出力データを標準化しません。つまり、すべての品目定義は、認識用には常に使用されますが、出力用には使用されない場合があります。これは、本番および QA の Oracle DataLens Server のみに適用され、開発サーバーでは非アクティブ・フラグは無視されます。デフォルトでは、Knowledge Studio は開発サーバーと同様に動作するため、非アクティブ・フラグは無視されます。

品目定義の非アクティブ・フラグは、「Phrases」タブの「Define Items」サブタブからのみ設定できます。

**非アクティブ・フラグの設定** データ・レンズが出力データを標準化するときに未完了の品目定義を使用しないようにするには、品目定義を右クリックし、次のいずれかのオプションを選択して非アクティブ・フラグを設定します。

### Set Inactive for Production

選択した品目定義のみ非アクティブに設定されます。

### Set Inactive this and all children for Production

選択した品目定義とそのすべての子が非アクティブに設定されます。

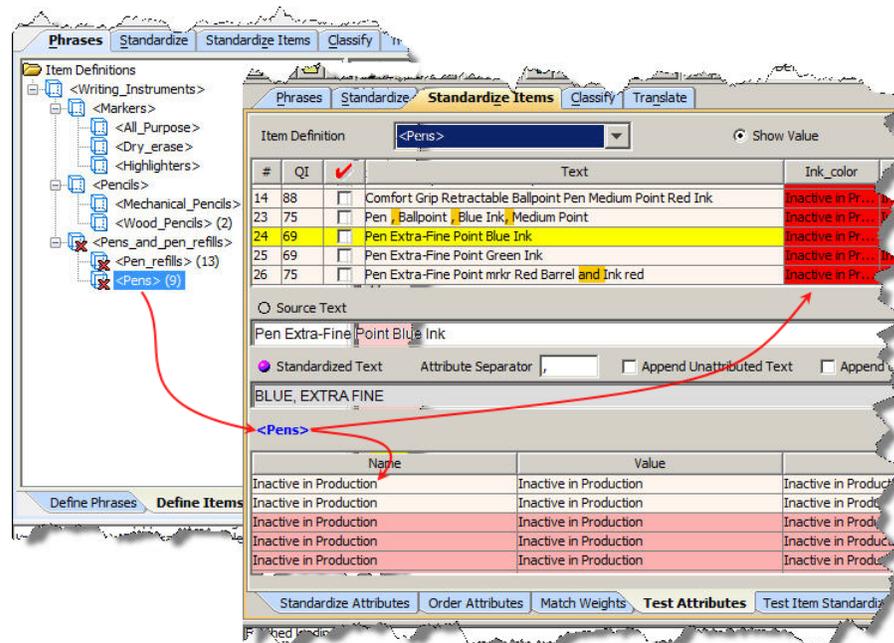
選択した品目定義のアイコンが変わるため、非アクティブに設定された品目定義を簡単に識別できます。親品目定義を展開して、子品目定義も非アクティブかどうかを確認してください。

子を含めて（子も非アクティブの場合）品目定義を再アクティブ化するには、品目定義を選択して右クリックし、「**Activate this and all children for Production**」を選択します。

**非アクティブな品目定義の出力データの表示** 非アクティブに設定された品目定義の標準化された出力データは、次のタブから表示できます。

- 「Standardize Items」タブの「Test Attributes」サブタブ
- 「Standardize Items」タブの「Test Item Standardization」サブタブ
- 「Translate」タブの「Test Translated Attributes」サブタブ
- 「Translate」タブの「Test Item Translation」サブタブ

いずれかのサブタブで、「View」メニューから「View as Production」を選択します。次の例に示すように、非アクティブに設定された品目定義の出力データが表示され、「Inactive in Production」とマークされます。



**注意：** 非アクティブ化された品目定義が品目定義のあいまい表現に含まれる場合は、続行する前にあいまい表現を削除してください。

## 「Item Definition List」 ペイン

このペインをアクティブ化するには、品目定義をダブルクリックするか、または品目定義を選択してコンテキスト依存メニューから「**Edit Item Definition**」オプションを選択します。「**Item Definition List**」ペインに表示されるアイコンおよびラベルは、品目定義の構成要素です。このペインで、選択した品目定義に対してスコアリング属性を設定できます。

品目定義に関する次の用語が使用されます。

### 属性

属性とは、品目の特性です。たとえば、ペン先サイズ、本体色およびインク色はすべて、ボールペンの属性です。属性値は、「細字」、「黒インク」、「赤インク」などです。

### 属性の重要度

オブジェクトを識別する際に、すべての属性が同じ役割を果たすわけではありません。たとえば、オブジェクト名「ボールペン」はオブジェクトを正しく定義するのに必要ですが、本体色は必ずしも必要ではありません。品目定義では、次の5つの方法で属性の重要度を定義します。

- **Required:** この属性は、オブジェクトを定義するために必ず表示される必要があります。
- **Scoring:** この属性は必須ではありませんが、スコアリング・アルゴリズムを使用して品目の定義を支援します。
- **Optional:** この属性は、品目を定義する際に必須ではなく、スコアリング・アルゴリズムにも関与しません。オプション属性では、結果のあいまいさは解消されません。一般的に、品目定義で用語がスコアリング属性または必須属性で使用されると、同じデータ・レンジ内のすべての品目定義でその用語がスコアリング属性または必須属性で使用されます。
- **Disallowed:** 品目が定義される際に表示されません。
- **Unused:** この属性は使用されません。これは、品目定義で属性を使用するかどうか、またはその使用方法が不明な場合に属性の一時記憶域として使用でき、品目定義が複雑な場合に役立ちます。

### Return to Item Definition List

「Item Definitions」ペインに戻ります。

### Drag to Create New...

これらのオプションを使用して、属性の重要度を設定します。

### Selected Item Definition

設定されている品目定義とすべての属性が表示され、属性の重要度の値も含まれます。

### Item Definition Score Table

この表はペインの左下隅にあり、選択したデータ行の品目定義メトリックが記録されます。「AM」は移入された必須属性とスコアリング属性の数、「Cnd」は一致した条件付き属性の数、「Spn」は必須属性とスコアリング属性によって範囲設定された合計文字数をそれぞれ示します。品目定義のあいまい表現が発生したデータ行がある場合は、この表をレビューするとあいまい表現の解決に役立ちます。

Item Definition	AM	Cnd	Spn	Status
Resistor	5	0	20	BEFORE

次のコンテキスト依存メニューは、「Item Definition List」ペインで使用できます。

#### Add Attribute

属性を追加します。詳細は、2-23 ページの「属性の追加」を参照してください。

#### Append AnyOf Below

選択したノードの下に任意の属性を追加するノード（ブール演算子）を追加します。

#### Append AllOf Below

選択したノードの下にすべての属性を追加するノード（ブール演算子）を追加します。

#### Append OneOf Below

選択したノードの下にいずれか1つの属性のみを追加するノード（ブール演算子）を追加します。

#### Delete Contents

1つの属性を削除するときに、確認メッセージを表示せずに、選択した属性を削除します。削除の確認を求めるメッセージが表示されないため、このオプションを使用する前に、正しい属性を選択していることを確認してください。

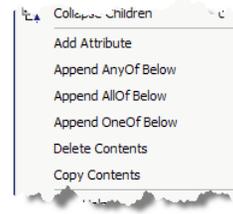
「Append」ノードのいずれかを削除する場合は、「Append」ノードとそれに含まれるすべてのノードを削除することを確認する削除確認メッセージが表示されます。

#### Copy Contents

別の属性にペーストするために、属性または階層の内容をコピーします。詳細は、2-23 ページの「品目定義間での属性のコピーとペースト」を参照してください。

#### Paste Replacement

別の属性にペーストするためにコピーした属性または階層の内容をペーストします。詳細は、2-23 ページの「品目定義間での属性のコピーとペースト」を参照してください。



次のコンテキスト依存メニューは、属性が選択された場合に使用できます。

#### Jump to Attribute Standardization

属性または属性に関連付けられた生成結果を選択するとアクティブになります。「Standardize Items」タブの「Standardize Attributes」サブタブに移動し、選択内容が表示されて編集できます。

#### Insert AnyOf Above

選択したノードの上に「AnyOf」ノードを挿入します。

#### Insert AllOf Above

選択したノードの上に「AllOf」ノードを挿入します。

#### Insert OneOf Above

選択したノードの上に「OneOf」ノードを挿入します。

#### Rename

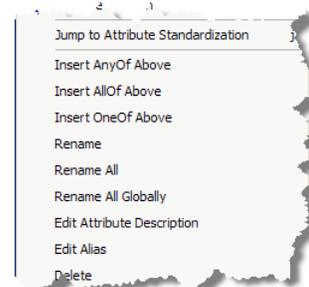
選択した属性の名前を、指定した名前に変更します。

#### Rename All

選択した属性および品目定義内で同じ名前のすべての属性の名前を、指定した名前に変更します。これは、アクティブなサブタブにのみ適用されます。

#### Rename All Globally

選択した属性およびデータ・レンズ全体で同じ名前のすべての属性の名前を、指定した名前に変更します。



### Edit Attribute Description

選択した属性摘要の名前を、指定した名前に変更します。

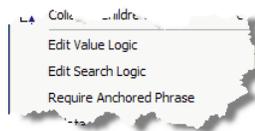
### Edit Alias

選択した属性別名の名前を、指定した名前に変更します。

次のコンテキスト依存メニューは、属性の下にあるフレーズが選択された場合に使用できます。

### Edit Value Logic

属性値ロジックを編集できます。詳細は、2-24 ページの「[属性値ロジックの変更](#)」を参照してください。



### Edit Search Logic

属性検索ロジックを編集できます。詳細は、2-26 ページの「[属性検索ロジックの変更](#)」を参照してください。このオプションは属性の用語またはフレーズ・ルールを選択した場合のみ使用可能になり、属性に含まれる属性を選択した場合は使用できません。

### Require Anchored Phrase

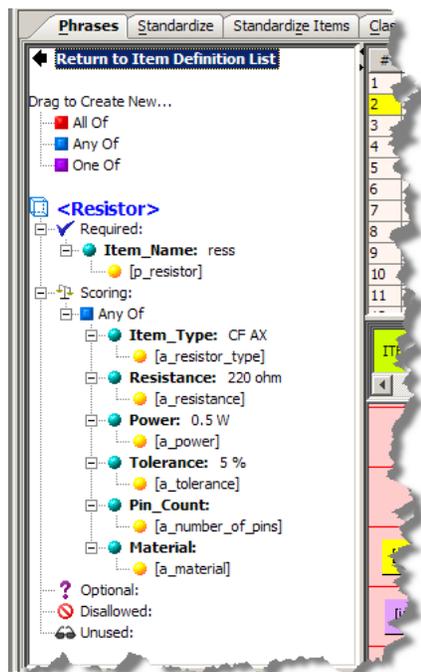
属性をフレーズ・ルールに固定する必要があるため、選択したフレーズには親を作成できません。

## ツールの使用

「Item Definition List」ペインを使用して標準化を行うと、変更内容は選択した品目定義のみに反映され、選択したフレーズまたは用語ルールに対するグローバル標準化は無視されます。

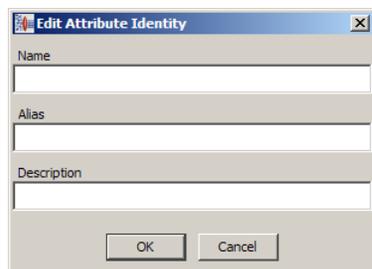
### 品目定義属性の定義

データを正しく認識するためにスコアリング属性は必須ではありませんが、品目のスコアリングに関与します。



用語を作成する際のベスト・プラクティスは、完全形式を入力することです。これによって、予測機能が強化されます。予測の詳細は、2-11 ページの「[ルール予測の使用](#)」を参照してください。

**属性の追加** 属性を任意の品目定義に追加するには、ノードを右クリックし、「Add Attribute」を選択します。



新規属性の名前を入力し（オプションで、別名および新規属性を説明するテキストも追加できます）、「OK」をクリックします。

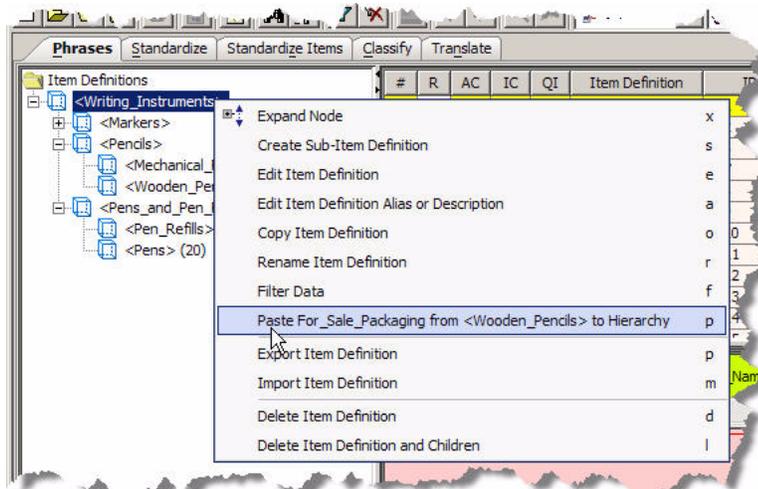
**品目定義間での属性のコピーとペースト** 品目定義間で属性をコピーすると、時間を節約できます。この項では、品目定義間で属性をコピーする方法について説明します。

次のサンプル・ステップでは、For\_Sale\_Packaging を <Wooden\_Pencils> 品目定義からデータ・レンズ内の他の品目定義にコピーします。

1. <Wooden\_Pencils> 品目定義をオープンします。
2. 「Optional」属性にある For\_Sale\_Packaging を選択します。
3. 右クリックして「Copy Contents」を選択します。



4. 「Return to Item Definition List」をクリックします。
5. 最上位レベル品目定義 <Writing\_Instruments> を右クリックし、「Paste For\_Sale\_Packaging from <Wooden\_Pencils> to Hierarchy」をクリックします。



これによって、コピーした属性が、最上位レベル品目定義の下位にある階層全体にペーストされます。

品目定義を表示すると、For\_Sale\_Packaging ノードが品目定義に含まれています。属性がコピーされる時、属性の重要度はコピー元と同じままです。

**属性値ロジックの変更** データ・レンズによってデータが誤って認識される場合があります。たとえば、C と F は温度スケールの摂氏と華氏の略語として認識されます。ただし、C と F は、データ・レンズ内の他のルールで用語バリエーションである場合があります。値ロジックを使用すると、品目定義内の無効な温度値を除外できるため、無効な温度範囲を除外できます。

---

**注意：** 数値属性値に対して値ロジックを使用する場合は、単位換算を使用するようにフレーズが構成されている必要があります（詳細は、11-6 ページの「[フレーズに対する単位換算の作成](#)」を参照）。

---

属性の評価に使用するロジックを変更するには、属性を右クリックして「**Edit Value Logic**」を選択します。

The screenshot shows a dialog box titled "Value Logic for Ink\_Type: [a\_ink\_type]". It is divided into two main sections: "Type of Value" and "Value Expression".

- Type of Value:** Contains three radio buttons: "Units with Conversion", "Numeric Only", and "Text". The "Text" option is selected. Below the radio buttons is a dropdown menu.
- Value Expression:** Contains three radio buttons: "Comparison:", "Range:", and "List:".
  - Comparison:** Selected. Shows "\$text" followed by an equals sign (=) in a dropdown, and "0.0" in a text box.
  - Range:** Shows "\$text from" followed by "0.0" in a text box, "to", and "0.0" in another text box. Below are two checked checkboxes labeled "Inclusive".
  - List:** Shows "\$text in" followed by a text box containing "0.0".
- Below the "List" section is a checked checkbox labeled "Case Sensitive".
- User-defined:** A radio button that is not selected, followed by a text box containing "\$text eq \"0.0\"".
- None:** A radio button that is not selected.

At the bottom of the dialog are "OK" and "Cancel" buttons.

属性のデフォルト値が表示されます。属性に応じて、「**Type of Value**」セクションが編集できない場合があります。

「**Value Expression**」セクションを使用して、属性値を評価する方法を変更できます。次の評価方法から1つのみを選択できます。

#### Comparison

定義した比較範囲内に生成結果に関連付けられた値がある場合、属性が移入されます。

#### Range

入力した範囲内に生成結果に関連付けられた値がある場合、属性が移入されます。

#### List

入力したリスト内に値があるかどうか評価されます。

#### User-defined

評価の正規表現を入力します。

#### None

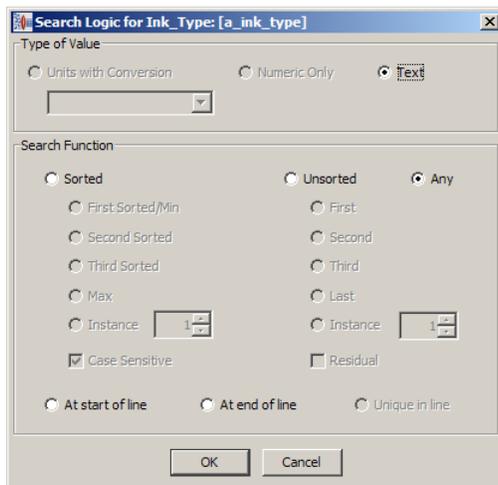
この属性に対して評価を実行しません。

---

**注意：** 数値属性値に対して値ロジックを使用する場合は、最初に、単位換算に対してフレーズが構成されている必要があります。詳細は、11-6 ページの「[単位換算タイプの削除](#)」を参照してください。

---

**属性検索ロジックの変更** 検索時に属性の評価に使用するロジックを変更するには、属性を右クリックして「Edit Search Logic」を選択します。



属性のデフォルト値が表示されます。属性に応じて、「Type of Value」セクションが編集できない場合があります。

「Search Function」セクションを使用して、検索時に属性値を評価する方法を変更できます。次の検索方法から1つのみを選択できます。

#### Sorted

数値がソートされ、該当する値が選択されます。結果値は、「Sorted」リスト内の指定のインスタンス（1番目、2番目、3番目、N番目など）です。

#### Unsorted

データ行に表示されている順序に基づいて、値が選択されます。

#### Any

すべての値が検索されます。

#### At start of line

値がデータ行の冒頭にある場合のみ、属性が移入されます。

#### At end of line

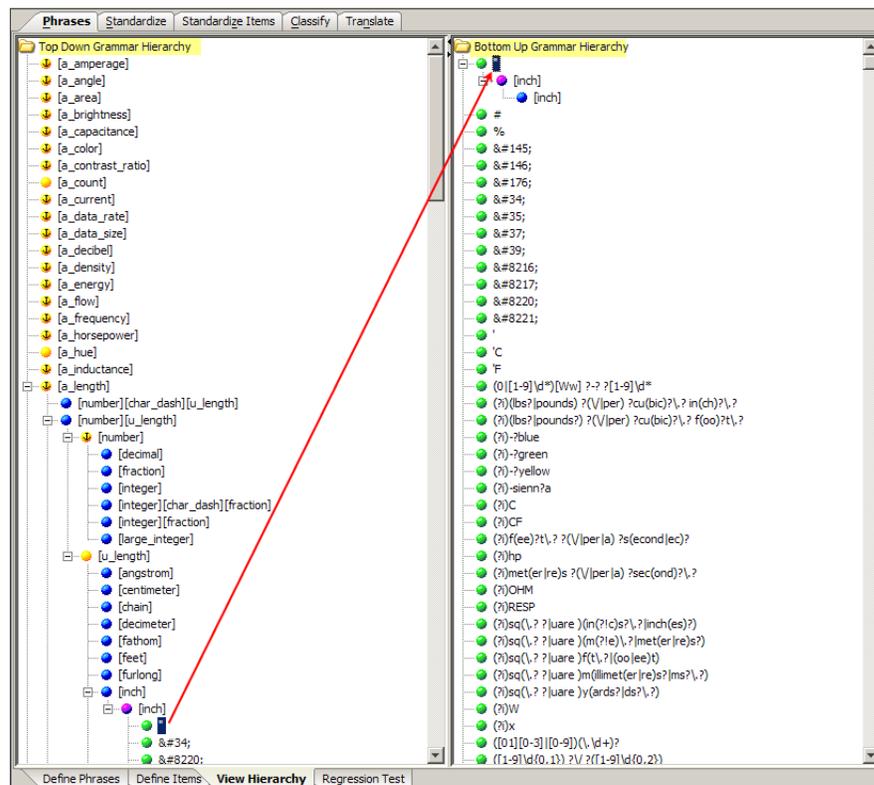
値がデータ行の末尾にある場合のみ、属性が移入されます。

#### Unique in line

この機能は非推奨です。

## 「View Hierarchy」 サブタブ

「View Hierarchy」サブタブを使用して、用語とフレーズの生成結果間の関係をレビューできます。



## 「Top Down Grammar Hierarchy」 および 「Bottom Up Grammar Hierarchy」 ペイン

これらのペインには、左側のトップ・ダウン階層にフレーズの生成結果が表示され、右側のボトム・アップ階層に関連する用語の生成結果が表示されます。いずれかのペインで生成結果を選択でき、反対側のペインで最上位レベルの生成結果に達すると、前の図に示すように生成結果が展開されます。

「Top Down Grammar Hierarchy」または「Bottom Up Grammar Hierarchy」ペインのコンテキスト依存メニューには、次のメニュー・オプションが含まれます。

### Show Hierarchy From Here

選択した親ノードから下位の階層構造全体を展開します。



### Show Hierarchy in Other Tree

選択内容が反対側の階層ツリーで検索されて自動的に展開され、ノード全体を表示できます。

### Jump to Rule for Editing

フレーズ、用語または生成結果を選択した場合のみアクティブになり、「Define Phrases」サブタブがアクティブ化され、選択内容が表示されて編集できます。

## 「Regression Testing」 サブタブ

回帰テストの目的は、データ・レンズの保守を行った際に、以前に定義したフレーズおよび用語と比較して、あいまい表現や問題が発生していないことを確認することです。これは重要なステップで、完成したデータ・レンズを微調整した場合のみ実行してください。

回帰セットには、このセットとともに作成されたサンプル・ファイルが連結されています。広範囲のデータをカバーするサンプル・ファイルを選択または作成すると、変更の結果を最もよく検証できます。通常、回帰ベースは大規模なファイルで、数千の行が含まれるのが一般的です。

---

**注意：** サンプル・データ・ファイルの最大行数は 5000 行で、回帰ベース (5000 行かどうかに関係なく) がデータ・セット全体の多様性を完全に網羅している必要があります。

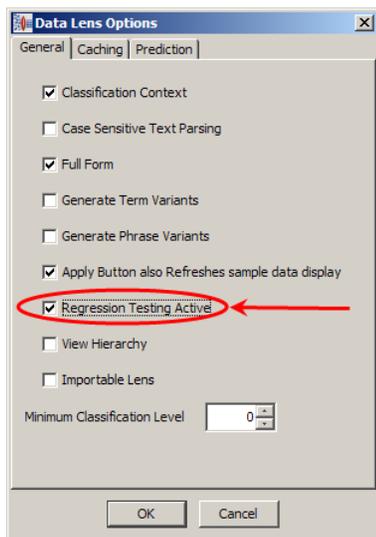
---

回帰テストは「Standardize Items」タブ、「Classify」タブおよび「Standardize」タブで最も有用ですが、「Phrases」タブでの有用性は限定的です。

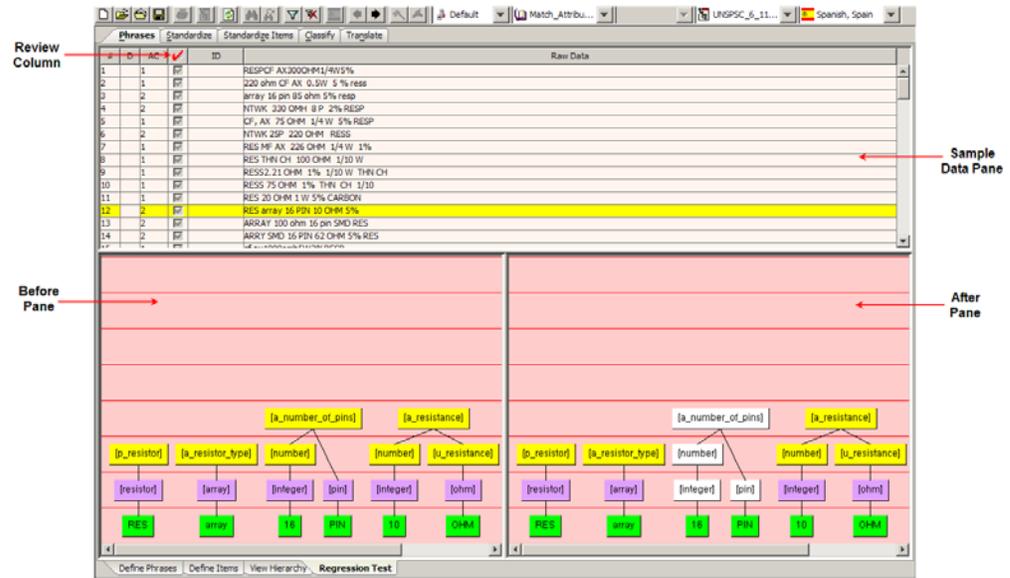
### 回帰テストがアクティブであることの確認

作業するデータ・レンズをオープンし、回帰テストを実行するサンプル・データ・ファイルを選択します。

現在のタブおよびすべてのタブで「Regression Test」サブタブをアクティブ化するには、「Data Lens」メニューから「Data Lens Options」をクリックします。「Regression Testing Active」チェック・ボックスを選択し、「OK」をクリックします。



「Regression Test」サブタブには、回帰テスト結果のルールがグラフィック表示されます。



## 「Before」 ペイン

このペインには、回帰テスト前のデータ・レンズでの定義に従って、「Sample Data Table」ペインで選択したデータ行のフレーズ・ルールがグラフィック表示されます。

## 「After」 ペイン

このペインには、「Sample Data Table」ペインで選択したデータ行のフレーズ・ルールがグラフィック表示され、レビューできるように差異が表示されます。

「Before」 ペインおよび「After」 ペインにデータが表示されない場合は、サンプル・データが初期化されていないため、回帰ベースは存在しません。回帰ベースの初期化の詳細は、2-29 ページの「回帰ベースの作成と更新」を参照してください。

## 「Review」 列

「Review」 列（赤色のチェック・マーク）は、新規または変更されたデータ行を示し、この行のテキストをレビューする必要があります。「Current Text」列の情報が正しく、有効な推移として変更を受け入れる場合は、このチェック・ボックスを選択して、変更が回帰テストに組み込まれるようにします。

## 回帰ベースの作成と更新

回帰ベースの作成におけるベスト・プラクティスは、サンプル・データを1つのファイルに結合することです。回帰テストは、完成したデータ・レンズに対して実行するのが最も効果的です。詳細は、8-12 ページの「サンプル・ファイルの結合」を参照してください。

ファイルの結合によってデータが削除されることはなく、選択したサンプル・ファイルが新規の大規模なファイルに単純に結合されます。

次に、回帰ベースのサンプル・データ・ファイルに対して1つの変更を加え、回帰セットを確認して適切に更新します。複数の変更を加えると回帰の結果が読みにくくなり、エラーを見落としたり修正が困難になる可能性があります。

回帰ベースを作成するには、「File」メニューの「Create New Regression Base」オプションを選択し、回帰テストに使用するサンプル・データ・ファイルを選択します。これによって、回帰ベースが初期化され、「After」ペインに結果が表示されます。

回帰ベースは、「File」メニューの「Update Regression Base」オプションを使用して、レビュー済および受入済のテキスト行（前述の「Review」列）で更新できます。

---

---

**注意：** サンプル・データをレビューまたは受け入れた場合のみ、回帰ベースを初期化または更新してください。

---

---

---

## データの標準化

データの標準化の目的は、データを一貫した明瞭なものにすることです。一貫性とは、共通の用語と形式を使用して関連するデータを識別できるように、出力の信頼性を確保することです。明瞭性とは、データ保守プロセスに関与していなくてもデータを簡単に理解できることです。

データを（データ・レンズ全体で）グローバルに標準化することも、詳細なレベルで標準化を実施することもできます。この章では、データをグローバルに標準化する方法について説明します。品目定義による標準化の詳細は、[第4章「品目定義の標準化」](#)を参照してください。

標準化は、出力の表示方法を制御します。データ・セット内の用語、フレーズまたは特定の属性の出力方法を選択できます。たとえば、データには複数の種類の「highlighter」という語が含まれている可能性があります。「highlighter」は、データ内で「hi-lighter」や「hi-liters」のように出現することがあります。highlighter の出力ルールを標準化することで、入力データに「hi-lighter」や「hi-liters」で表されているかどうかに関係なく、「highlighters」となるこの語の標準化を選択できます。

## 「Standardize」タブ

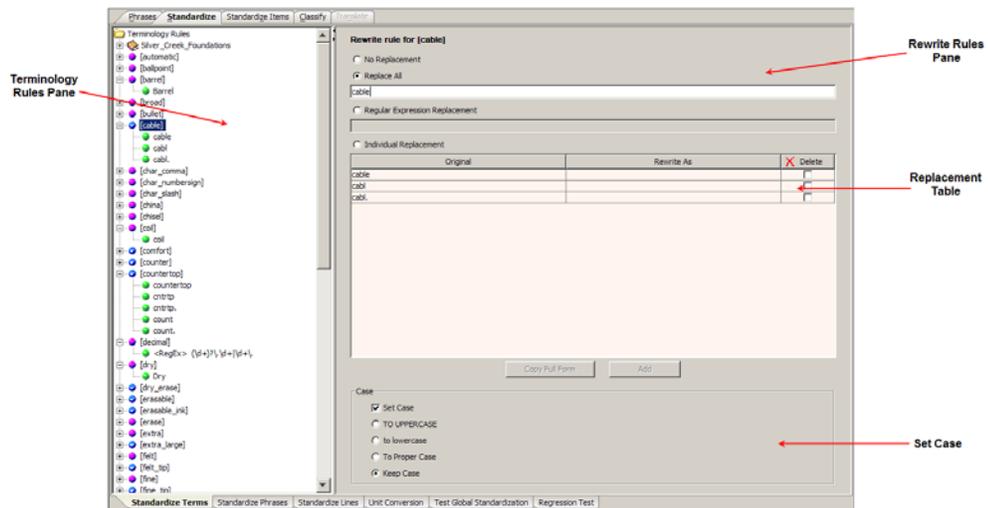
「Standardize」タブおよび関連するサブタブには、データをグローバルに標準化の際に必要なすべての機能が用意されています。

## 「Standardize Terms」サブタブ

「Standardize Terms」サブタブは、初めて「Standardize」タブを選択したときにデフォルトで表示されます。次のことができます。

- 用語ルールに割り当てられている用語バリエーションすべてに対する置換方法の割当て
- 大 / 小文字の設定
- 置換値の簡単なコピーとペースト

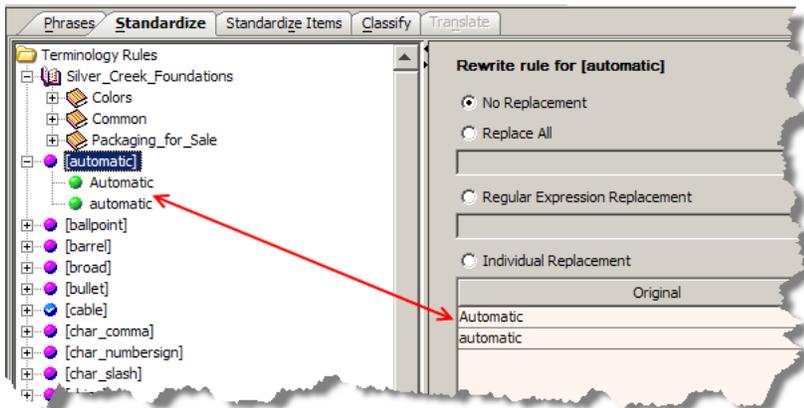
図 3-1



## 「Terminology Rules」ペイン

「Terminology Rules」ペインには、レンズで発生する用語ルールの完全なセットが表示されます。このセットには、作成したルールとスマート用語集を使用してインポートされたルールが含まれます。

用語ルールをダブルクリックするか、横にあるプラス (+) 記号でルールを展開すると、その用語バリエーションが表示されます。これらのバリエーションは「Rewrite Rule」ペインにも表示されます。



シングルクリックではルールは展開されませんが、用語バリエントが「**Rewrite Rule**」ペインに表示されます。

## 「Rewrite Rules」ペイン

このペインのメイン領域では、データ・レンズでテキスト表現を置換して用語を標準化するために使用する、次の4種類のいずれかの方法を選択できます。

### No Replacement

解析時にバリエントは変更されません。用語ルールの用語バリエントに対する標準化は行われません。

### Replace All

すべてのバリエントは、フィールドに入力したテキストに置換されます。「**Define Phrases**」サブタブを使用して用語ルールに後から追加した用語は、自動的に標準化されます。

### Regular Expression

すべてのバリエントは、フィールドに入力した正規表現に置換されます。

### Individual Replacements

置換表の「**Rewrite As**」列をクリックすると、置換テキストを入力できるフィールドがアクティブ化されます。各バリエントを個別に変更して異なるテキストを指定できます。

「**Add**」ボタンをクリックと、用語バリエントをさらに追加できます。新しい用語バリエントを追加するには、新しい行の「**Original**」フィールドをクリックして指定のテキストを入力し、「**Rewrite As**」フィールドに置換テキストを入力します。

「**Rewrite As**」フィールドを消去するには、「**Delete**」チェック・ボックスを選択します。

このペインのコンテキスト依存メニューは、**置換表**でのみアクティブ化され、メニュー・オプションは置換の選択によって異なります。可能性のあるメニュー・オプションは次のとおりです。

#### Make this the Replace All value

このオプションを使用して、「**Replace All**」方式の値を移入します。

Make this the Replace All value
Copy table to Clipboard
Paste Clipboard to table

#### Copy table to Clipboard

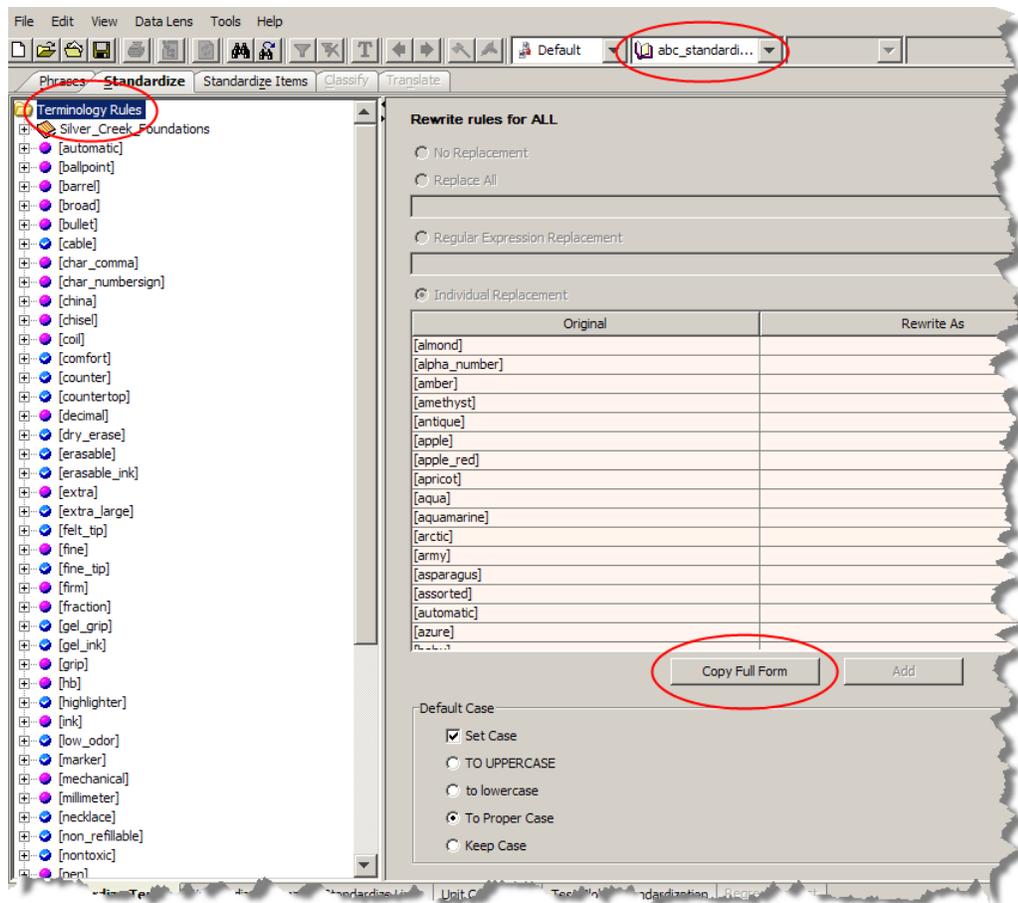
このオプションを使用して、置換表の内容を別のアプリケーションにペーストできるように、クリップボードにコピーします。たとえば、Excel スプレッドシートにペーストする場合です。

#### Paste Clipboard to table

別のアプリケーションでテキスト・リライト・ルールを作成し、そのルールを置換表に個別に入力するのではなく直接ペーストする場合は、このオプションを使用します。

## 「Copy Full Form」ボタン

「**Copy Full Form**」ボタンを使用して、データ・レンズで用語ルールに対して定義した完全形式に基づいて、リライト・ルールを簡単に作成できます。このボタンは、同じ名前のペインで「**Terminology Rules**」フォルダ全体が選択されている場合にアクティブになります。



現在使用されている「Standardization Type」のデフォルトの大 / 小文字は、「Default Case」セクションで設定され、「Copy Full Form」をクリックすることでリライト・ルールに適用されます。

たとえば、「To Proper Case」設定を使用することは、各用語ルールの完全形式（略語でない）がデフォルトの出力であることを意味します。略語「med」が元のデータにある場合も、完全形式の「medium」が標準化出力形式です。

### 「Default Case」セクション

データの大 / 小文字が区別される場合は、「Set Case」チェック・ボックスをクリックし、適切なチェック・ボックスを使用して大 / 小文字の変更方法を選択することで、用語バリエントを標準化する必要があります。

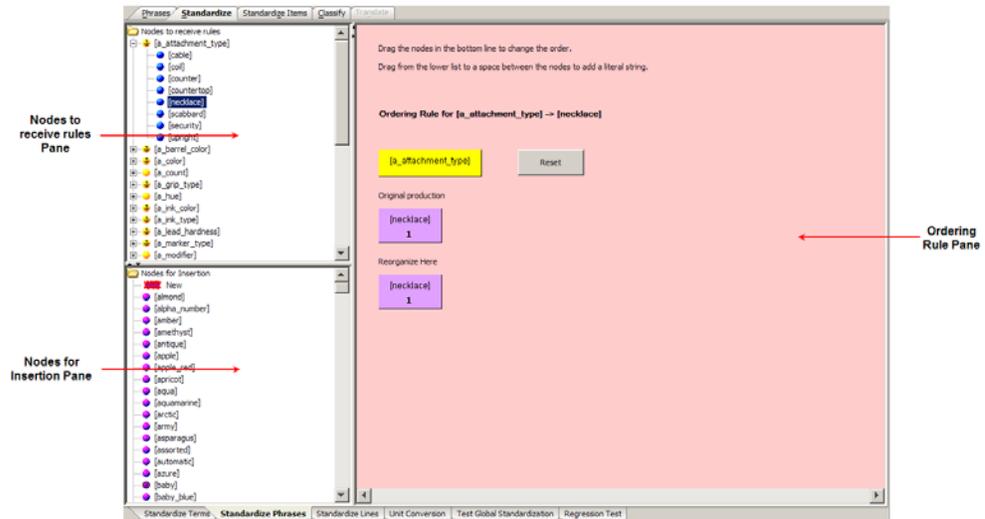
- TO UPPERCASE
- to lowercase
- To Proper Case
- Keep Case

大 / 小文字タイプの各オプションは、用語バリエントとして表示され、標準化されます。デフォルトは「Keep Case」で、大 / 小文字は変更されません。

## 「Standardize Phrases」サブタブ

「Standardize Phrases」サブタブでは、次のことができます。

- フレーズ内の個々の用語とフレーズ要素の順序とその要素の組合せ方法を指定します。
- 新しい用語ルールを追加してフレーズを修正します。



### 「Nodes to Receive Rules」ペイン

「Nodes to Receive Rules」ペインには、データ・レンズ内の全フレーズのリストが表示されます。順序変更するノードを選択でき、選択すると「Ordering Rule」ペインにそのノードが表示されます。

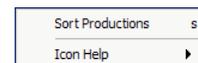
「Ordering Rules」ペインで変更を追加すると、フレーズに対する変更がこのペインに反映され、フレーズ・アイコンの色または形状が変化します。

### Sort Productions

このペインのコンテキスト依存メニューは、フレーズを選択するとアクティブ化されます。

#### Sort Productions

このメニュー・オプションを使用して、「Ordering Rule」ペインを使用して適用した変更内容を、選択されたノードのすべての生成結果をソートして表示します。



### 「Nodes for Insertion」ペイン

「Nodes for Insertion」ペインには、データ・レンズ内の全用語のリストが表示されます。既存の用語ルールを追加して生成結果を標準化する場合は、「Nodes for Insertion」ペインから「Ordering Rule」ペインにノードをドラッグします。

## 「Ordering Rule」 ペイン

「Ordering Rule」 ペインでは、フレーズを操作してデータ・レンズ内の各ノードの正確な順序を作成できます。このペインには、2つの異なる機能表現があり、ルールの変更、生成結果用語の結合、またはこのペインの生成結果の削除を実行できます。

「Nodes to receive rules」 ペインでフレーズまたはフレーズの生成結果を展開すると、「Ordering Rule」 ペインに適切な並替え機能が表示されます。

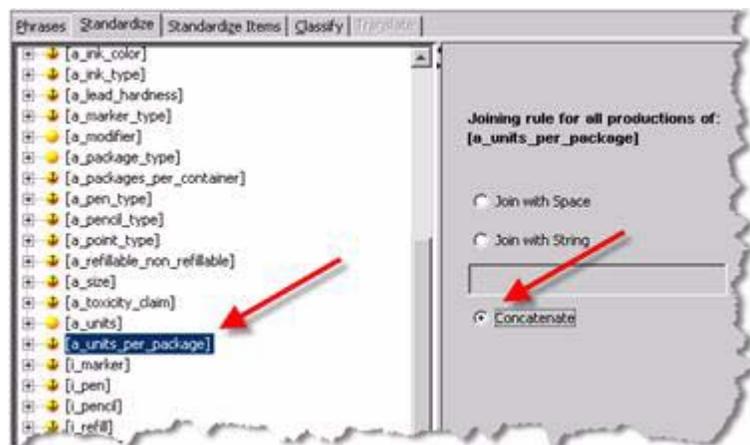
「Ordering Rule」 ペインには、生成結果に対して実行できる3つの操作があります。

- 用語の追加
- 用語の削除
- 用語の順序の変更

### 用語の結合

ある用語を別の用語と結合する機能によって、データを迅速に標準化し、用語を簡潔にできます。空白以外の文字で結合することが役に立つ特定のフレーズが存在することがあります。たとえば、「2/Package」と表示されるように、「2 / Package」のスラッシュの前後の空白を削除できます。

「Ordering Rule」 ペインで機能を変更するノードを選択し、用語を結合（または連結）して、ノードの生成結果に新しいルールを反映できます。



3つの連結オプションは、次のとおりです。

#### Join with Space

このフレーズの各生成結果のそれぞれの用語は、標準化された出力では空白で区切られます。

#### Join with String

このフレーズに関連付けられたすべての複合語生成結果の用語は、指定した文字列（文字）で結合されます。

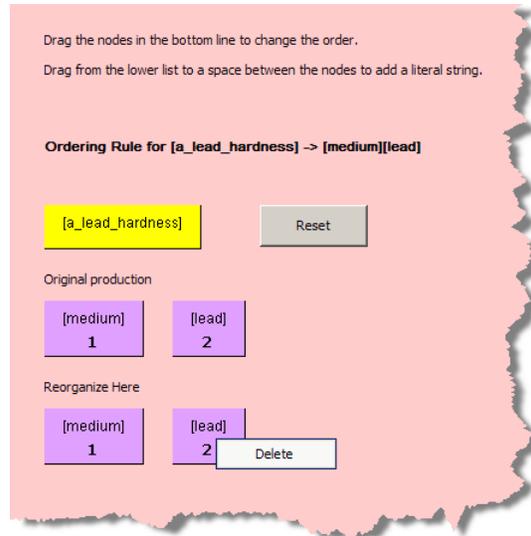
#### Concatenate

このフレーズの各複合語生成結果の用語は、空白なしで連結されます。

## 用語の削除または順序変更

アクティブな機能で用語を削除または順序変更する場合は、元の生成結果順序が表示され、移動または削除可能な生成結果がその下に表示されます。

用語を右クリックして「Delete」をクリックすることで、このペインで用語を削除できます。



フレーズの生成結果が「Reorganize Here」セクションにドラッグされ、新しいフレーズの標準化ルール（順序変更）が作成されます。

「Reset」ボタンは、フレーズを元の生成結果に戻し、削除も含めた変更内容をすべて無効にします。

---

**注意：**「Ordering Rule」ペインで行ったグローバル標準化のすべての変更を表示するには、「Test Global Standardization」サブタブを使用します。

---

## 「Standardize Lines」サブタブ

「Standardize Lines」サブタブはすでに有効利用されておらず、将来のリリースで廃止される予定です。



# 「Test Global Standardization」 サブタブ

「Test Global Standardization」サブタブでは、サンプル・データに適用したグローバル標準化ルールをレビューできます。

#	QI	Len	Initial Text	Standardized Text
1	100	47	Mechanical Pencil .5MM Emerald Barrel	Mechanical Pencil 0.5 millimeter Emerald Barrel
2	100	44	Pencil .5MM Green Barrel Automatic	Pencil 0.5 millimeter Green Barrel Automatic
3	100	49	Balpoint Pen Medium Point Black Barrel Black Ink	Balpoint Pen Medium Point Black Barrel Black Ink
4	100	54	Balpoint Pen Refill Fine Point Blue Ink 2 / Pack	Balpoint Pen Refill Fine Point Blue Ink 2 per package
5	100	51	Balpoint Pen Refill Med Pt Black Ink 2 / Ink	Balpoint Pen Refill Med Pt Black Ink 2 per package
6	100	56	Balpoint Pen Refill Med, Point Black, ink 2 / Pack	Balpoint Pen Refill Med, Point black ink 2 per package
7	100	40	Balpoint Pen Refill Point Mdm Black Ink	Balpoint Pen Refill Point Mdm Black Ink
8	100	55	Balpoint Pen Refill mdm point .7mm ink Black	Balpoint Pen Refill mdm point 0.7 millimeter ink Black
9	100	35	China Mkr Green Ink 12 / Box	China Mkr Green Ink 12 per package
10	100	21	China Markers red ink	China Markers red ink
11	100	23	China Markers White ink	China Markers White ink
12	100	20	China Mkr yellow ink	China Mkr yellow ink
13	100	38	Chisel Point Highlighter Turquoise ink	Chisel Point Highlighter Turquoise ink
14	100	60	Comfort Grip Retractable Balpoint Pen medium Point Blue Ink	Comfort Grip Retractable Balpoint Pen medium Point Blue Ink
15	100	59	Comfort Grip Retractable Balpoint Pen Medium Point Red Ink	Comfort Grip Retractable Balpoint Pen Medium Point Red Ink
16	100	44	Counter top pen refill Medium Point black Ink	Counter top pen refill Medium Point black Ink
17	100	56	Dry-Erase Chisel Point Markers Black Ink 2 / Set	Dry - Erase Chisel Point Markers Black Ink 2 per package
18	100	53	Dry-Erase Chisel Tip Markers Assorted 4 / Set	Dry - Erase Chisel Tip Markers Assorted 4 per package
19	100	42	Security Pen Refills Medium Point Blue Ink	Security Pen Refills Medium Point Blue Ink
20	100	41	Counter Pen Refills Medium Point Blue Ink	Counter Pen Refills Medium Point Blue Ink
21	100	40	Dry-Erase Marker Bullet Point Blue Ink	Dry - Erase Marker Bullet Point Blue Ink
22	100	38	Dry-Erase Marker Fine Point Blue Ink	Dry - Erase Marker Fine Point Blue Ink
23	100	36	Dry Erase Markers Chisel Tip Red Ink	Dry Erase Markers Chisel Tip Red Ink
24	100	52	Retractable Balpoint Pen Non-Refillable Black Ink	Retractable Balpoint Pen Non - Refillable Black Ink
25	100	60	Erasable Ink Balpoint Pen Med Point Blue Barrel,Blue Ink	Erasable Ink Balpoint Pen Med Point Blue Barrel / Blue Ink
26	100	62	Erase Tip Pen Eraser, Blue Point, Turquoise Ink, 12 / Box, 75 mm, 12.5 mm	Erase Tip Pen Eraser, Blue Point, Turquoise Ink, 12 / Box, 75 mm, 12.5 mm

Source Text (editable - press Enter when done)

Mechanical Pencil .5MM Emerald Barrel

Formatted Text

Mechanical Pencil .5MM Emerald Barrel

Standardized Text

Mechanical Pencil 0.5 millimeter Emerald Barrel

Item Attributes

Mechanical\_Pencils: Product\_Name=Mechanical, Type=Pencil, Size=0.5 millimeter, Barrel\_color=Emerald Barrel, For\_Sale\_Packaging=msd,

Tagged Standardized Text

[a\_pencil\_type] (Mechanical) [l\_pencil] (Pencil) [a\_size] (0.5 millimeter) [a\_barrel\_color] (Emerald Barrel)

Standardize Terms | Standardize Phrases | Standardize Lines | Unit Conversion | **Test Global Standardization** | Regression Test

## サンプル・データ表

この表には、元のデータと標準化された後の同じデータが表示されます。5つの列は、左から右に次の内容を示しています。

### 行番号 (#)

データ行に割り当てられている一意の番号。

### 品質指標 (QI)

0 から 100 までの数値で、行が標準化された程度を表しています。

### 長さ (Len)

そのデータ行に含まれている文字数。

### 赤色のチェック・マーク

レビューしてマーク付けしたデータで、該当するデータ行でチェック・ボックスを選択します。

### 元テキスト

データ・レンズで解析された、標準化されていない元のデータ行。

### Standardized Text

元のデータ行に対する標準化書式。

データが含まれている各列は、その列ヘッダーをクリックすることで表を昇順と降順の両方でソートできます。列ヘッダーを1回クリックすると、表がソートされ、選択した列の品目別に昇順のアルファベット順で表示されます。同じ列ヘッダーを2回クリックすると、表が再度ソートされ、降順のアルファベット順で表示されます。

## サブタブのデータ・フィールド

サンプル・データ表内の行のいずれかを選択すると、次の情報が表の下のデータ・フィールドに表示されます。

### 「Standardized Text」 フィールド

元のデータに対する標準化バージョン。

### 「Tagged Standardized Text」 フィールド

初期データに対して完全にタグが設定され、標準化されたバージョン。

### 「Formatted Text」 フィールド

ソース書式設定ルールを適用した結果。

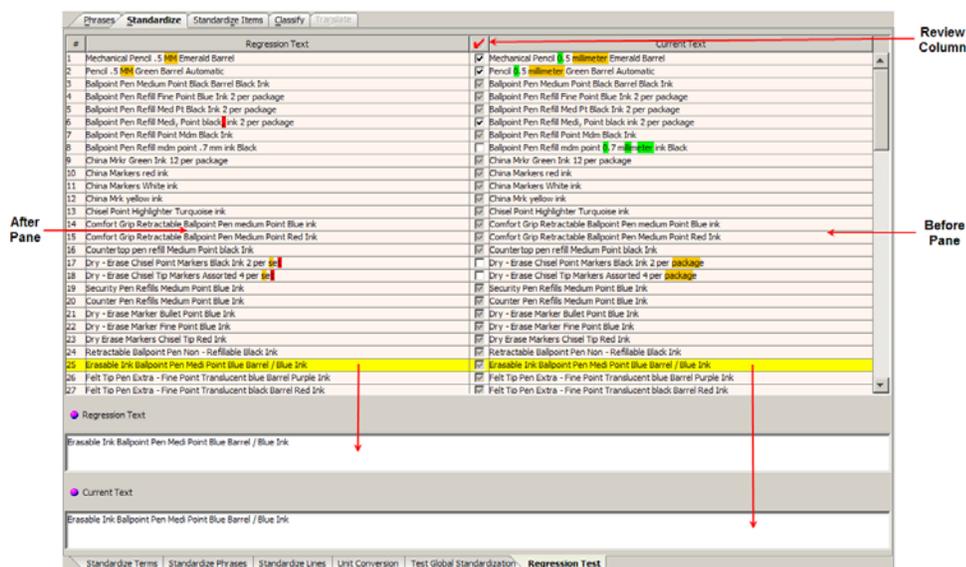
### 「Source Text」 フィールド

元のデータ。このフィールドは編集可能で、[Enter] キーを押すと、データ・レンズ上で即時に結果をレビューできます。

## 「Regression Test」 サブタブ

回帰テストはデータ標準化の重要な部分で、データ出力が予想どおりであることを確認できます。

タブがアクティブでない場合は、「Regression Testing Active」データ・レンズ・オプションを設定します。詳細は、8-2 ページの「データ・レンズ・オプションの設定」を参照してください。



回帰テストには、サンプル・データの「前」の状態と「後」の状態を示す2つのペインがあります。

## 「Before」 ペイン

このペインには、回帰テスト前にデータ・レンズに定義されていたルールに基づいて標準化されたデータが表示されます。このペインの選択したデータ行に表示されるテキストは、「Current Text」フィールドにも表示されます。

## 「After」 ペイン

このペインには、データ・レンズに定義されたルールに基づいて標準化されたテキストが表示されます。このペインの選択したデータ行に表示されるテキストは、「Regression Text」フィールドにも表示されます。

「Before」ペインおよび「After」ペインにデータが表示されない場合は、サンプル・データが初期化されていないため、回帰ベースは存在しません。回帰ベースの初期化の詳細は、3-11ページの「回帰ベースの作成と更新」を参照してください。

「Before」ペインまたは「After」ペインの色付けされたテキストは、次の内容を示します。

### 赤

削除されたデータ。

### 緑

追加されたデータ。すべてのテキストは問題についてレビューする必要があるため、左右のペインを目視で比較します。

### オレンジ

この用語には該当する標準化が適用されているため、回帰データと現在のデータの両方が色付けされます。

## 「Review」列

「Review」列（赤色のチェック・マーク）は、新規または変更されたデータ行を示し、この行のテキストをレビューする必要があります。「Current Text」列の情報が正しく、有効な推移として変更を受け入れる場合は、このチェック・ボックスを選択して、データが回帰テストに含まれるようにします。

## 回帰ベースの作成と更新

回帰ベースの作成におけるベスト・プラクティスは、サンプル・データを1つのファイルに結合することです。詳細は、8-12ページの「サンプル・ファイルの結合」を参照してください。

ファイルの結合によってデータが削除されることはなく、選択したサンプル・ファイルが新規の大規模なファイルに単純に結合されます。

次に、回帰ベースのサンプル・データ・ファイルに対して1つの変更を加え、回帰セットを確認して適切に更新します。複数の変更を加えると回帰の結果が読みにくくなり、エラーを見落としたり修正が困難になる可能性があります。

回帰ベースを作成するには、「File」メニューの「Create New Regression Base」オプションを選択し、回帰テストに使用するサンプル・データ・ファイルを選択します。これによって、回帰ベースが初期化され、「After」ペインに結果が表示されます。

回帰ベースは、「File」メニューの「Update Regression Base」オプションを使用して、レビュー済および受入済のテキスト行（前述の「Review」列）で更新できます。

---

---

**注意：** サンプル・データをレビューまたは受け入れた場合のみ、回帰ベースを初期化または更新してください。

---

---

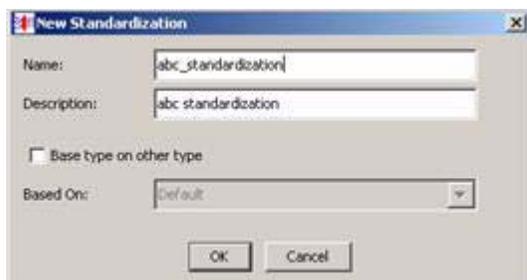
## 標準化のタイプ

Enterprise DQ for Product では、デフォルトの標準化タイプが提供されます。したがって、標準化タイプを追加せずに、即時に標準化ルールを追加できます。

### 標準化タイプの作成

データ・レンズ全体で使用する独自の標準化スキーマを作成できます。標準化タイプは、標準化のすべてのタブおよびサブタブで使用されます。

1. 「Data Lens」メニューから「Standardization Types...」を選択します。
2. 「Add New」ボタンをクリックします。



3. 新しい標準化タイプの作成に必要な情報を入力し、選択オプションとして「Standardization Types」リストに追加します。
4. 作成した標準化タイプがすでにあり、同じ標準化の新しいバージョンでそのナレッジを再利用する場合は、「Base classification on other type」チェック・ボックスを選択し、次に「Based On:」リストから適切な分類タイプを選択します。

---

**注意：** 他の標準化タイプがない場合、このチェック・ボックスはアクティブになりません。

---

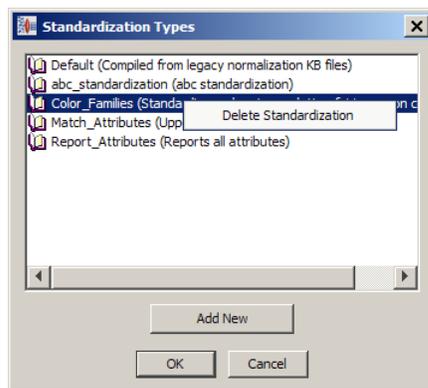
5. 「OK」をクリックします。

ツールバーの標準化タイプ・リストを使用して、編集する新しい標準化タイプを選択できます。ツールバー・リストから標準化タイプを選択すると、選択したタイプに関連付けられている標準化ルールが「Standardize」タブに表示されます。「Phrases」タブの「Define Phrases」サブタブでのフレーズと用語の保守によって、すべての標準化タイプが更新されます。データ・レンズを保存すると、標準化タイプに加えられたすべての変更が保存されます。

## 標準化タイプの削除

標準化タイプは、必要に応じて削除できます。

1. 最新のデータ・レンズ・バージョンをチェックインしたことを確認します。
2. 「Data Lens」メニューから「Standardization Types...」を選択します。
3. 削除する標準化タイプを選択して右クリックします。



4. 「Delete Standardization」をクリックします。  
削除確認ダイアログが表示されます。
5. 選択した標準化タイプを削除する場合は「OK」をクリックし、削除しない場合は「Cancel」をクリックします。
6. 「OK」をクリックします。

標準化タイプの使用の詳細は、11-5 ページの「[単位の標準化タイプ](#)」を参照してください。



---

## 品目定義の標準化

個々の品目定義の標準化は、前の章で説明されているグローバル標準化の割当てとは異なります。品目定義の標準化では、特定の品目定義のコンテキスト内でデータを標準化する方法および標準化を1つの品目定義にのみ適用する方法を選択します。

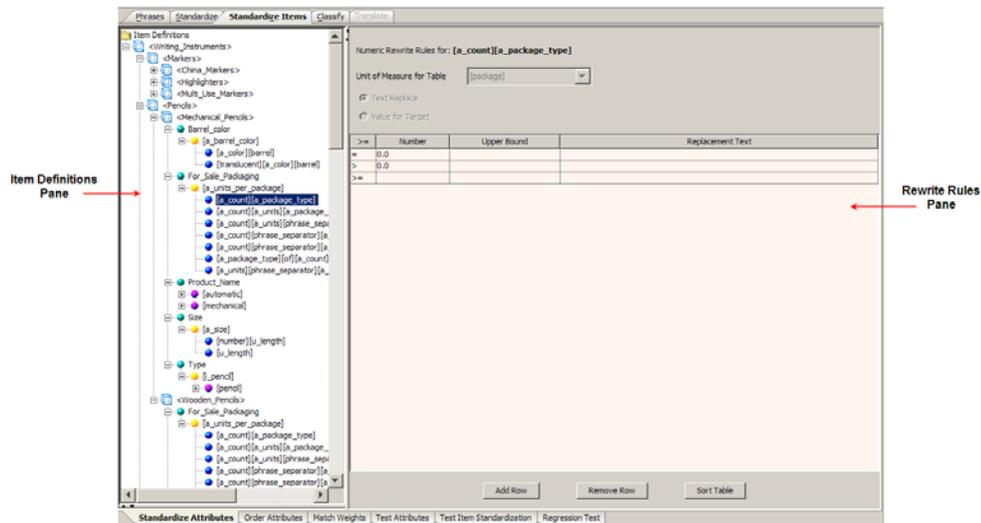
たとえば、ボールペンの色は鉛筆の色と区別して標準化できます。ボールペンの色属性に関するルールを作成すると、出力データには、ボールペンのみに関連した色を含めることができます。鉛筆の色のみに関連する出力のために、鉛筆に関する別の色ルールを定義することもできます。これらのルールは、すべての品目定義の色をグローバルに標準化するのではなく、ボールペンおよび鉛筆の品目定義のみを標準化しますが、任意のグローバル標準化によって上書きされます。

## 「Standardize Items」 タブ

「Standardize Item」タブおよび関連するサブタブには、品目定義レベルでデータを標準化する際に必要なすべての機能が用意されています。

## 「Standardize Attributes」 サブタブ

「Standardize Attributes」サブタブは、初めて「Standardize Items」タブを選択したときにデフォルトで表示されます。



「Standardize Attributes」サブタブには、複数の個別機能表現があり、数値によるルールのリライト、ルールのグローバルな自動リライト、またはフレーズの生成結果の順序変更を実行できます。

## 「Item Definitions」 ペイン

「Item Definitions」ペインには、データ・レンズ内のすべての品目定義が表示されます。品目定義、属性、フレーズまたは用語を選択できます。

親品目定義のすべての階層が表示されますが、すべての子品目定義の元となる属性は自動的に表示されるわけではありません。表示して標準化するすべての属性構造を展開するには、品目定義をダブルクリックする必要があります。代わりに、「Expand Node」オプションを使用することもできます。ただし、親品目定義に属性が含まれている場合（青色ではなく赤色の正方形アイコンで識別）は、その品目定義と子品目定義が自動的に展開され、すべて表示されます。

---

**注意：** 属性のみが含まれている親品目定義（赤色の正方形アイコンで表示）の属性は、他のすべての属性と同じ方法で標準化できます。

---

品目定義を展開すると、その属性は品目定義階層に追加されてメモリ内にあるため、表示されたままになります。大規模な品目定義階層全体の表示には大量のメモリが必要となるため、この動作によって「Item Definition」ペインへのデータ移入が最小メモリで実現します。

「Standardize Attributes」サブタブの動作は、このペインでの選択に従って異なります。たとえば、用語を選択すると、「Nodes for Insertion」ペインおよび「Rewrite Rules」ペインの表示が「Ordering Rule」ペインの表示に変更されます。

このペインのコンテキスト依存メニューには、次のメニュー・オプションが表示される可能性があります。

**Expand Node**

現在選択している階層構造を展開するには、このオプションを使用します。

**Copy to...**

11-11 ページの「品目定義標準化の共有」を参照してください。

**Copy with Children to...**

11-11 ページの「品目定義標準化の共有」を参照してください。

**Copy to another Item Definition...**

11-11 ページの「品目定義標準化の共有」を参照してください。

**Jump to Standardization**

「Standardize」タブの「Standardize Phrases」サブタブに移動して、グローバル標準化ルールを設定するには、このオプションを使用します。

	Expand Node	x
	Copy to...	o
	Copy with Children to...	c
	Copy to another Item Definition...	i
	Icon Help	▶

	Expand Node	x
	Jump to Standardization	j
	Icon Help	▶

**「Rewrite Rules」ペイン**

「Rewrite Rules」ペインを使用すると、「Item Definitions」ペインで選択したタイプに基づいて4つの異なる方法が提供されるため、標準化ルールの作成が容易になります。

**数値リライト・ルール**

「Rewrite Rules」ペインの**数値リライト・ルール**表現を使用すると、標準単位などの数値に基づいて数値ベースの品目定義属性を標準化できます。

デフォルトの「Numeric Rewrite Rules」ペインはテキスト置換です。

Numeric Rewrite Rules for: [a\_size]

Unit of Measure for Table: [millimeter]

Text Replace  
 Value for Target

>=	Number	Upper Bound	Replacement Text
=	0.0		Empty
>	0.0	< 0.5	Drafting
>=	0.5	< 0.7	Office
>=	0.7		

「Unit of Measure for Table」リストを使用すると、リストの下にある表内で標準化する数値を選択できます。このリストには、データ・レンズのすべての単位が移入されます。

**ボタン** ボタンは次のように使用されます。

**Add Row**

表に一度に1行ずつ行が追加されます。

### Remove Row

アクティブな行または下の行が削除されます。一部の行は削除できません。

---

**注意：** 削除の確認プロンプトは表示されず、行を元に戻すことはできません。

---

### Sort Table

表がソートされます。

**オプション「Text Replace」** オプションおよび **「Value for Target」** オプションを使用すると、選択したルールの置換方法（テキストまたは値による置換）を選択できます。

**テキスト置換表** テキスト置換表の各列は、次のように動作します。

>=

下限を示し、編集できません。

### Number

標準化テキストに置換するための数値で、「Upper Bound」列のエントリに基づいて自動的に移入されます。

### Upper Bound

この置換ルールを特定の**数値**（上限）で停止するための数値。前の行に従って、この列のフィールドは自動的に移入されるか、または新しいデータを入力できます。

### Replacement Text

特定の**数値**を置換するテキストを入力できます。

### 値置換表

値置換表の動作は、テキスト置換表と類似しています。

Numeric Rewrite Rules for: [a\_size]

Unit of Measure for Table: [millimeter]

Text Replace

Value for Target

>=	Number	Upper Bound	Target	UOM Spelling
=	0.0		[millimeter]	
>	0.0	< 0.5	[millimeter]	
>=	0.5	< 0.7	[millimeter]	
>=	0.7		[millimeter]	

Add Row Remove Row Sort Table

最初の3つの列は、前述の説明のとおり動作します。「Replacement Text」列は「Target」列および「UOM Spelling」列に置き換えられ、テキストのかわりに値を指定できます。

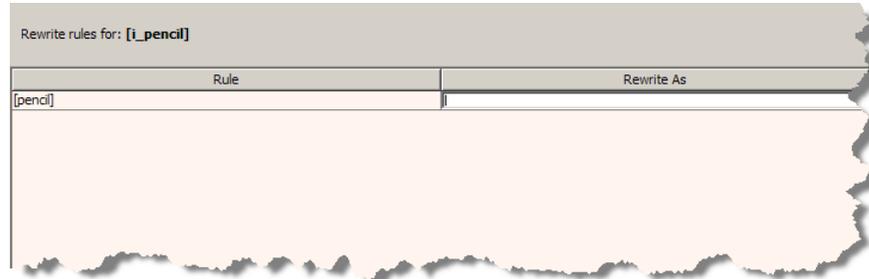
「Target」列は、選択が可能な単位のリストです。

「UOM Spelling」列には、選択したターゲットに対する出力データのスペルを入力できます。

特定のフレーズに関連付けられている値置換表を削除する場合は、そのフレーズを右クリックして「Reset Table」をクリックします。

### 簡単なリライト・ルール

この簡素化された「Rewrite Rules」ペインでは、「Rewrite As」フィールドに置換テキストを入力して、選択したルールに新しいテキストを定義できます。



### 自動のリライト・ルール

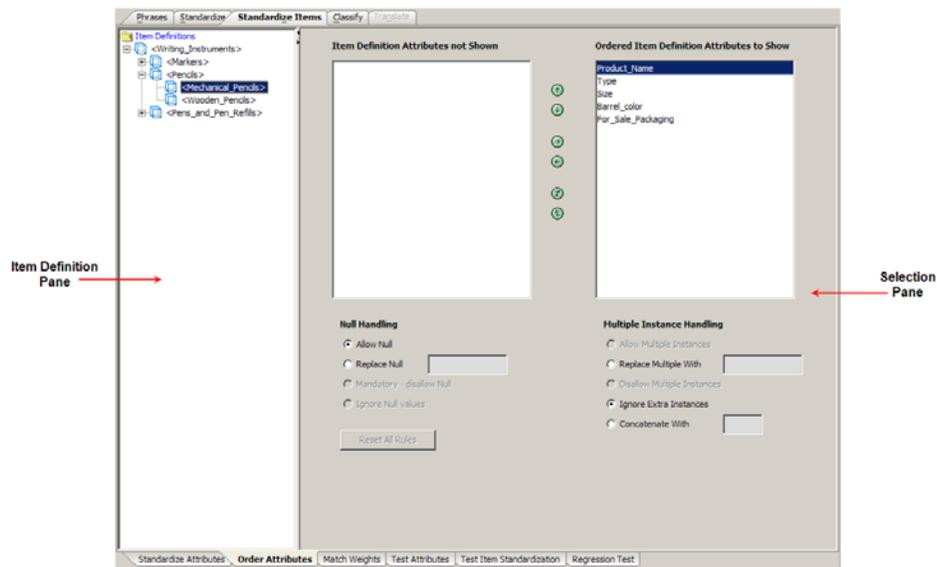
「Item Definitions」ペインから用語を選択すると、「Numeric Rewrite Rule」ペインが自動の「Rewrite Rules」ペインに変更されて、選択した用語にグローバル標準化ルールを設定できます。この機能は、前の章で説明されている「Standardize」タブの「Standardize Phrases」サブタブと同じです。

### フレーズの生成結果の順序変更

「Item Definitions」ペインからフレーズの生成結果を選択すると、「Numeric Rewrite Rule」ペインが「Ordering Rule」ペインに変更され、フレーズの順序変更または生成結果の追加を実行できます。この機能は、前の章で説明されている「Standardize」タブの「Standardize Phrases」サブタブと同じです。

## 「Order Attributes」 サブタブ

「Order Attributes」 サブタブでは、品目定義属性が出力データに表示される順序を定義できます。



### 「Item Definition」 ペイン

このペインでは、データ・レンズ内の特定の品目定義のみ、またはすべての品目定義を選択できます。

---

**注意：** このセクションのアクティブな選択肢は、選択した品目定義に基づいています。

---

### 「Selection」 ペイン

「Selection」 ペインでは、属性を「Ordered Item Definition Attributes to Show」リストから「Item Definition Attributes not Shown」リストに移動して、照合（表示）する品目定義属性を選択できます。「Ordered Item Definition Attributes to Show」リストのすべての属性は、照合に使用され、標準化出力に表示されます。

属性をリスト間で移動するには、右矢印と左矢印を使用するか、属性をダブルクリックします。照合と出力の表示順序を変更するには、上矢印と下矢印を使用します。

「Reset All Rules」 ボタンがアクティブになるのは、「Item Definitions」 フォルダが選択されている場合のみです。これは、データ・レンズ内のすべての品目定義に対する「Null」設定および「Multiple Instance Handling」設定をすべてリセットする際に使用できます。

#### Null Handling

「Null Handling」 セクションでは、NULL 属性値の処理方法および「Selection」 ペインにリストされているすべての属性に対する適用方法を選択できます。NULL 値の処理の選択肢は次のとおりです。

##### Allow Null

空の属性が許可されます。

##### Replace Null

空の属性は、このフィールドに入力したテキストに置換されます。

**Mandatory - disallow Null**

データ・レンズ内で空の値は許可されません。

**Ignore Null values**

空の値は無視されます。

**Multiple Instance Handling**

「Multiple Instance Handling」セクションでは、複数の属性値の処理方法を選択できます。処理の選択肢は次のとおりです。

**Allow Multiple Instances**

1つの属性に対して複数のインスタンスが許可されます。

**Replace Multiple With**

1つの属性に対する複数のインスタンスは、このフィールドに入力したテキストに置換されます。

**Disallow Multiple Instances**

この機能は非推奨です。

**Ignore Extra Instances**

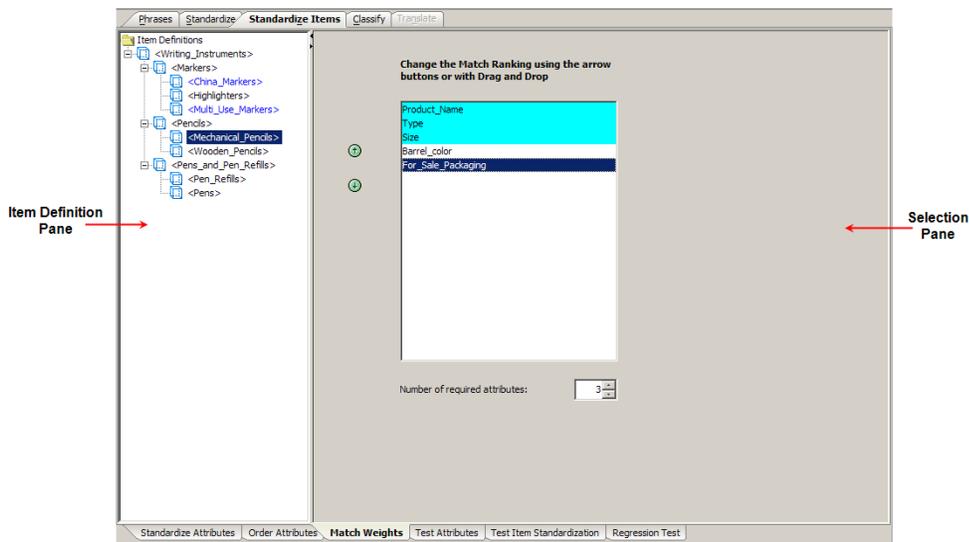
最初に出現する属性が出力で選択され、後続のすべての属性値は無視されます。

**Concatenate With**

1つの属性に対する複数の一意のインスタンスが特定の文字列とともに連結されます。さらに、この操作では属性値が、その性質に応じて数値順またはアルファベット順にソートされます。

## 「Match Weights」 サブタブ

「Match Weights」サブタブでは、照合する属性の数、および選択した属性を特定の品目定義に対して照合する順序を設定できます。このサブタブは、照合タイプが 4-14 ページの「Match Type」で説明されているように作成されるまでアクティブになりません。



### 「Item Definition」 ペイン

このペインでは、照合ランク付けを変更する品目定義を選択できます。

### 「Selection」 ペイン

「Selection」ペインでは、リストに表示された品目定義属性から照合する属性を選択できます。これは、「Number of required attributes」リスト内の数値を増減することで自動的に選択されます。この選択はリスト内の最初の品目から開始され、順番に下方向に選択されます。

選択した属性の照合ランク付けを変更するには、上矢印と下矢印を使用し、照合に使用する最も重要な属性をリストの最上部に移動します。これらの矢印は、リストから属性を 1 つ選択するとアクティブになります。

各品目定義で照合する属性を識別すると、照合タイプが完了します。照合品目を作成するためのこれらの属性は、照合する DSA 内で使用されます。

---

**注意：** データの照合は、Application Studio、Knowledge Studio および Governance Studio を使用して構成された複合プロセスです。

---

### 重複照合

品目定義に対して重複および類似データが識別および照合されます。重複データの識別がデータ・セット内とデータ・セット間の両方で必要な理由は多数あります。重複の識別の問題は、レコードが同じ書式でないために、標準の文字列の比較方法では検出できない場合があります。たとえば、「Ballpoint Pen Refill Med Pt Black Ink 2 / Pk」と「Ballpoint Pen Refill Medi, Point black, ink 2 / Pack」は「Pack」のスペルが異なりますが、同じ品目を指している場合があります。

照合の作成の基盤はデータ・レンズ内にありますが、データの実際の照合は DSA 内で実行されます。照合機能は、属性値の比較に基づいています。複数のデータ行の照合属性値が同等の場合、これらの品目は一致として識別されます。

ユーザーが開発したユースケースに従って、部品番号、製造業者およびブランドを照合できます。この場合は、照合プロセスに使用するこれら 3 つの属性を選択します。

別のユースケースの例では、形状、適応および機能に基づいて品目を照合する必要があります。この照合は、長さ、幅、高さ、色および材料に基づいて定義できます。

次の内容を理解することが重要です。

- 照合には、標準化タイプと照合タイプの両方が必要です。
- 照合は、特定の品目定義内に出現する属性に対して実行されます。
- 照合には、NULL 属性および複数属性の処理に対する特定の設定が必要です。

## 「Test Attributes」 サブタブ

「Test Attributes」 サブタブでは、サンプル・データに適用された品目定義別に属性標準化ルールをレビューして、結果を検証できます。

The screenshot displays the 'Test Attributes' sub-tab with the following components:

- Item Definition:** A table with columns: #, QI, Text, Product\_Name, Type, Size, Barrel\_color, For\_Sale\_Packa... The first row is highlighted in yellow.
- Sample Data Table:** A table with columns: #, QI, Text, Product\_Name, Type, Size, Barrel\_color, For\_Sale\_Packa... The first row is highlighted in yellow.
- Source Text:** A text input field containing 'Mechanical Pencil .5MM Emerald Barrel'.
- Standardized Text:** A text output field containing 'MechanicalPencilOfficeEmerald Barrel'.
- Standardized Attribute Table:** A table with columns: Name, Value, Text, Numeric Value, Unit of Measure. The first row is highlighted in red.

## 「Item Definition」 セクション

データ・レンズ内のすべての品目定義が選択用にリストされます。

「Show Value」 オプションおよび「Show Text」 オプションを使用すると、データの表示方法を変更できます。単位を表示するには、「Show Text」を選択します。

データ・レンズ内の任意の品目定義の属性をテストするには、リストから別の品目定義を選択します。このタブのすべてのデータが変更され、選択した新しい品目定義に関連するデータが表示されます。

デフォルトでは、アクティブな品目定義のデータのみが表示されます。非アクティブな品目定義データの表示の詳細は、2-18 ページの「[アクティブまたは非アクティブな品目定義](#)」を参照してください。

## サンプル・データ表

この表には、元のデータと標準化された後の同じデータが表示されます。各列は、左から右に次の内容を示しています。

### 行番号 (#)

データ行に割り当てられている一意の番号。

### 品質指標 (QI)

0 から 100 までの数値で、行の品目定義が標準化された程度を表しています。

### 赤色のチェック・マーク

レビューしてマーク付けしたデータで、該当するデータ行でチェック・ボックスを選択します。

### 元テキスト

データ・レンズで解析された元のデータ。

### 残りの各列

残りの各列は、各品目定義の属性に依存するため、これらの列は変化します。

データが含まれている各列は、そのタイトルをクリックすることで表を昇順と降順の両方でソートできます。列ヘッダーを 1 回クリックすると、表がソートされ、選択した列の品目別に昇順のアルファベット順で表示されます。同じ列ヘッダーを 2 回クリックすると、表が再度ソートされ、降順のアルファベット順で表示されます。

サンプル・データ表内の行のいずれかを選択すると、**標準化属性表**で選択した行の品目定義情報が表示されます。

## ソース・フィールド

このフィールドには、元のデータが表示されます。このフィールドは編集可能で、**[Enter]** キーを押すと、データ・レンズ上で即時に結果をレビューできます。

## 「Standardized Text」 セクション

このフィールドには元のデータの標準化バージョンが表示されますが、編集はできません。

「Attribute Separator」 フィールドでは、属性値にデリミタを指定できます。

「Append Unattributed Text」 チェック・ボックスを使用すると、品目定義の一部ではないが認識されているデータを挿入できます。

同様に、「Append Unparsed Text」 チェック・ボックスでは、未解析のテキストが摘要に挿入されます。

## 標準化属性表

この表には、選択した品目定義のすべての標準化属性および「Sample Data」 ペインで選択した行が表示されます。このフィールドの表示方法には、「Standardized Text」 セクションから選択した属性標準化が反映されます。

## 「Test Item Standardization」 サブタブ

「Test Item Standardization」サブタブでは、作成してサンプル・データに適用した品目定義の標準化ルールをレビューして、結果を検証できます。

#	QI	Len	Initial Text	Standardized Text
1	100	42	Mechanical Pencil .5MM Emerald Barrel	Mechanical, Pencil, Office, Emerald Barrel
2	100	39	Pencil .5MM Green Barrel Automatic	Automatic, Pencil, Office, Green Barrel
3	88	53	Balpoint Pen Medium Point Black Barrel Black Ink	Pen, Balpoint, Black Ink, Black Barrel, Medium Point
4	88	51	Balpoint Pen Refill Fine Point Blue Ink 2 / Pack	Balpoint, Blue Ink, Fine Point, Pen, 2 per package
5	88	48	Balpoint Pen Refill Med Pt Black Ink 2 / Pack	Balpoint, Black Ink, Med Pt, Pen, 2 per package
6	88	53	Balpoint Pen Refill Medi. Point black, ink 2 / Pack	Balpoint, black ink, Medi. Point, Pen, 2 per package
7	88	36	Balpoint Pen Refill Point Mdm Black Ink	Balpoint, Black Ink, Point Mdm, Pen
8	100	52	Balpoint Pen Refill mdm point .7mm ink Black	Balpoint, ink Black, 0.7 millimeter, mdm point, Pen
9	100	38	China Mkr Green Ink 12 / Box	China, Mkr, Green Ink, 12 per package
10	100	23	China Markers red ink	China, Markers, red ink
11	100	25	China Markers White ink	China, Markers, White ink
12	100	22	China Mkr yellow ink	China, Mkr, yellow ink
13	100	40	Chisel Point Highlighter Turquoise ink	Highlighter, Chisel Point, Turquoise ink
14	88	64	Comfort Grip Retractable Balpoint Pen medium Point Blue Ink	Pen, Retractable Balpoint, Blue Ink, Comfort Grip, medium Point
15	88	63	Comfort Grip Retractable Balpoint Pen Medium Point Red Ink	Pen, Retractable Balpoint, Red Ink, Comfort Grip, Medium Point
16	75	40	Counter top pen refill Medium Point black Ink	black Ink, Medium Point, Counter top, pen
17	100	60	Dry-Erase Chisel Point Markers Black Ink 2 / Set	Markers, Dry - Erase, Black Ink, Chisel Point, 2 per package
18	100	57	Dry-Erase Chisel Tip Markers Assorted 4 / Set	Markers, Dry - Erase, Assorted, Chisel Tip, 4 per package
19	75	37	Security Pen Refills Medium Point Blue Ink	Blue Ink, Medium Point, Security, Pen
20	75	36	Counter Pen Refills Medium Point Blue Ink	Blue Ink, Medium Point, Counter, Pen
21	100	43	Dry-Erase Marker Bullet Point Blue Ink	Marker, Dry - Erase, Blue Ink, Bullet Point
22	100	41	Dry-Erase Marker Fine Point Blue Ink	Marker, Dry - Erase, Blue Ink, Fine Point
23	100	39	Dry Erase Markers Chisel Tip Red Ink	Markers, Dry Erase, Red Ink, Chisel Tip
24	88	55	Retractable Balpoint Pen Non-Refillable Black Ink	Pen, Retractable Balpoint, Black Ink, Non - Refillable

Source Text (editable - press Enter when done)  
Balpoint Pen Medium Point Black Barrel Black Ink

Standardized Text  
Pen, Balpoint, Black Ink, Black Barrel, Medium Point

Item Definition  
Pens: Product\_Name=Pen, Type=Balpoint, Ink\_Color=Black Ink, Size=null, Ink\_Type=null, Barrel\_Color=Black Barrel, Grip\_Type=null, Point\_Type=Medium Point,

デフォルトでは、アクティブな品目定義のデータのみが表示されます。非アクティブな品目定義データの表示の詳細は、2-18 ページの「[アクティブまたは非アクティブな品目定義](#)」を参照してください。

## サンプル・データ表

この表の複数の列は、「Test Attributes」サブタブ（4-9 ページの「[Test Attributes](#)」サブタブ）を参照）の列と同じで、同じように動作します。異なる列は次のとおりです。

### 長さ (Len)

元のテキストの文字列の長さを示します。

### Standardized Text

標準化の元のテキストを示します。

### ソース・フィールド

このフィールドには、元のデータが表示されます。このフィールドは編集可能で、[Enter] キーを押すと、データ・レンズ上で即時に結果をレビューできます。

## 「Standardized Text」 セクション

このフィールドには元のデータの標準化バージョンが表示されますが、編集はできません。

「Attribute Separator」 フィールドには、属性間で使用するテキスト・セパレータを入力できます。

「Append Unattributed Text」 チェック・ボックスを選択すると、属性でないすべてのテキストが摘要に挿入されます。

同様に、「Append Unparsed Text」 チェック・ボックスでは、未解析のテキストが摘要に挿入されます。

## 「Item Attributes」 セクション

このフィールドには選択した品目定義のすべての属性が表示されますが、編集はできません。このフィールドの表示方法には、「Standardized Text」 セクションから選択した属性標準化が反映されます。

## 「Regression Test」 サブタブ

回帰テストはデータ標準化の重要な部分で、データ出力が予想どおりであることを確認できます。

品目の標準化回帰テストは、品目定義が設定されている行に対してのみ機能します。この行を認識できる品目定義が存在しない場合、回帰は表示されません。

タブがアクティブでない場合は、「Regression Testing Active」 データ・レンズ・オプションを設定します。詳細は、2-28 ページの「回帰テストがアクティブであることの確認」を参照してください。

#	AC	AD	Regression	Current	ID	Standardized Data
1	5/5	Mechanical_Pencils	<input type="checkbox"/>	Mechanical_Pencils	WI1	Mechanical Pencil 0.5 millimeter Emerald Barrel
2	5/5	Mechanical_Pencils	<input type="checkbox"/>	Mechanical_Pencils	WI2	Pencil 0.5 millimeter Green Barrel Automatic
3	8/8	Pens	<input checked="" type="checkbox"/>	Pens	WI6	Balpoint Pen Medium Point Black Barrel Black Ink
4	8/7	Pen_Refills	<input checked="" type="checkbox"/>	Pen_Refills	WI7	Balpoint Pen Refill Fine Point Blue Ink 2 per package
5	8/7	Pen_Refills	<input checked="" type="checkbox"/>	Pen_Refills	WI8	Balpoint Pen Refill Med Pt Black Ink 2 per package
6	8/7	Pen_Refills	<input checked="" type="checkbox"/>	Pen_Refills	WI9	Balpoint Pen Refill Med, Point black ink 2 per package
7	8/7	Pen_Refills	<input checked="" type="checkbox"/>	Pen_Refills	WI10	Balpoint Pen Refill Point Mdm Black Ink
8	8/7	Pen_Refills	<input checked="" type="checkbox"/>	Pen_Refills	WI11	Balpoint Pen Refill mdm point 0.7 millimeter ink Black
9	4/4	China_Markers	<input type="checkbox"/>	China_Markers	WI12	China Mkr Green Ink 12 per package
10	4/4	China_Markers	<input type="checkbox"/>	China_Markers	WI13	China Markers red ink
11	4/4	China_Markers	<input type="checkbox"/>	China_Markers	WI14	China Markers White ink
12	4/4	China_Markers	<input type="checkbox"/>	China_Markers	WI15	China Mkr yellow ink
13	4/4	Highlighters	<input type="checkbox"/>	Highlighters	WI16	Chisel Point Highlighter Turquoise ink
14	8/8	Pens	<input type="checkbox"/>	Pens	WI17	Comfort Grip Retractable Balpoint Pen medium Point Blue ink
15	8/8	Pens	<input type="checkbox"/>	Pens	WI18	Comfort Grip Retractable Balpoint Pen Medium Point Red Ink
16	8/7	Pen_Refills	<input type="checkbox"/>	Pen_Refills	WI19	Countertop pen refill Medium Point black Ink
17	6/6	Multi_Use_Markers	<input type="checkbox"/>	Multi_Use_Markers	WI20	Dry - Erase Chisel Point Markers Black Ink 2 per package
18	6/6	Multi_Use_Markers	<input type="checkbox"/>	Multi_Use_Markers	WI21	Dry - Erase Chisel Tip Markers Assorted 4 per package
19	8/7	Pen_Refills	<input type="checkbox"/>	Pen_Refills	WI22	Security Pen Refills Medium Point Blue Ink
20	8/7	Pen_Refills	<input type="checkbox"/>	Pen_Refills	WI23	Counter Pen Refills Medium Point Blue Ink
21	6/6	Multi_Use_Markers	<input type="checkbox"/>	Multi_Use_Markers	WI24	Dry - Erase Marker Bullet Point Blue Ink
22	6/6	Multi_Use_Markers	<input type="checkbox"/>	Multi_Use_Markers	WI25	Dry - Erase Marker Fine Point Blue Ink
23	6/6	Multi_Use_Markers	<input type="checkbox"/>	Multi_Use_Markers	WI26	Dry Erase Markers Chisel Tip Red Ink
24	8/8	Pens	<input type="checkbox"/>	Pens	WI27	Retractable Ballpoint Pen Non - Refillable Black Ink
25	8/8	Pens	<input type="checkbox"/>	Pens	WI28	Erasable Ink Ballpoint Pen Medi Point Blue Barrel / Blue Ink

Product Name	Type	Ink Color	Size	Ink Type	Barrel Color	Grip Type	Point Type
Pen	Balpoint	Black Ink			Black Barrel		Medium Point

Product Name	Type	Ink Color	Size	Ink Type	Barrel Color	Grip Type	Point Type
Pen	Balpoint	Black Ink			Black Barrel		Medium Point

## サンプル・データ表

この表には、サンプル・データと属性の回帰および現在のビューが表示されます。

**サンプル・データ表**にデータが表示されない場合は、サンプル・データが初期化されていないため、回帰ベースは存在しません。回帰ベースの初期化の詳細は、4-13 ページの「[回帰ベースの作成と更新](#)」を参照してください。

各列は、左から右に次の内容を示しています。

### #

サンプル・データの行に割り当てられている行番号。

### AC

属性件数には、回帰および現在の属性の属性件数が表示されます。

### AD

属性相違には、現在のファイルと回帰ファイル間の属性の相違件数が表示されます。

### Regression

回帰ファイルに基づいてデータ行に割り当てられている品目定義。

### 赤色のチェック・マーク

「**Review**」列（赤色のチェック・マーク）は、新規または変更されたデータ行を示し、この行のテキストをレビューする必要があります。「**Current Text**」列の情報が正しく、有効な推移として変更を受け入れる場合は、このチェック・ボックスを選択して、データが回帰テストに含まれるようにします。

回帰ベースは、「**File**」メニューの「**Update Regression Base**」オプションを使用して、レビュー済および受入済のテキスト行で更新します。

### Current

現在のファイルに基づいてデータ行に割り当てられている品目定義。

### ID

サンプル・データに割り当てられている一意の識別子で、サンプル・データの一部として含まれています。

### Standardized Data

これは、設定されている可能性があるグローバルな行順序の標準化に基づいています。

## 「Current Attribute」セクションおよび「Regression Attribute」セクション

これらのセクションには、**サンプル・データ表**で選択したデータ行の現在の属性と回帰属性が表示されます。

## 回帰ベースの作成と更新

回帰ベースの作成におけるベスト・プラクティスは、サンプル・データを1つのファイルに結合することです。詳細は、8-12 ページの「[サンプル・ファイルの結合](#)」を参照してください。

ファイルの結合によってデータが削除されることはなく、選択したサンプル・ファイルが新規の大規模なファイルに単純に結合されます。

次に、回帰ベースのサンプル・データ・ファイルに対して1つの変更を加え、回帰セットを確認して適切に更新します。複数の変更を加えると回帰の結果が読みにくくなり、エラーを見落としたり修正が困難になる可能性があります。

回帰ベースを作成するには、「**File**」メニューの「**Create New Regression Base**」オプションを選択し、回帰テストに使用するサンプル・データ・ファイルを選択します。これによって、回帰ベースが初期化され、「**After**」ペインに結果が表示されます。

回帰ベースは、「**File**」メニューの「**Update Regression Base**」オプションを使用して、レビュー済および受入済のテキスト行（前述の赤色チェック・マーク列）で更新できます。

---

---

**注意：** サンプル・データをレビューまたは受け入れた場合のみ、回帰ベースを初期化または更新してください。

---

---

## Match Type

Enterprise DQ for Product 照合機能を使用するには、データ・レンズ全体で使用する照合スキーマ（つまり照合タイプ）を作成する必要があります。1つ以上の照合タイプを作成し、ユースケースに従ってデータの照合方法を変更するために使用できます。

---

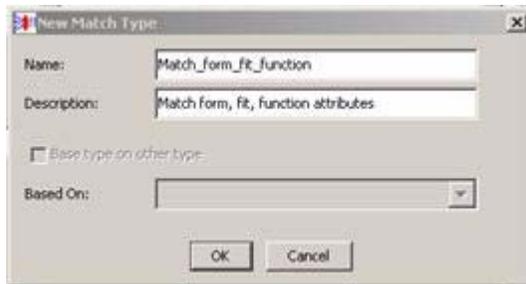
---

**注意：** 4-6 ページの「[「Order Attributes」サブタブ](#)」の説明に従って照合タイプを作成する前に、複数のインスタンスおよび NULL 値の処理方法を構成する必要があります。

---

---

この機能には、「Data Lens」メニューから「Match Types...」をクリックし、次に「Add New」ボタンをクリックしてアクセスします。



新しい照合タイプの作成に必要な情報を入力し、選択オプションとして「Match Types」リストに追加します。

「Match Weights」サブタブをアクティブにするには、照合タイプの作成と選択が必要です。

---

## データの分類

データを分類する主要な理由は、データを一般的な方法で把握するためです。データは、特定のキー、一意の属性またはそのデータを定義している SKU の知識がない場合でも、検索できます。多くの場合、製品に関するデータは、その製品のプロパティや使用方法を定義している特性の観点から検索します。たとえば、部品データの分類は、購買担当が、購買する必要がある部品を検索する際に役立ちます。これを実行するには、分類システムのいずれかのタイプに対して部品を定義または結びつけます。

分類システムは通常、将来の取得に備えてデータを分類できる階層構造です。広範な品目を包含する大規模な構造となる場合がありますが、類似品目を区別するための最小粒度を備えています。一方で、企業固有の複数の階層で分類システムを構成し、データの範囲を定義することもできます。

## 分類スキーマ

次の各項では、Oracle Enterprise Data Quality for Product Data で使用できる分類スキーマについて説明します。

### UNSPSC

製品とサービスに関する国際標準分類コード（UNSPSC）分類タイプは、電子商取引コード管理協会（ECCMA）を介して国連が開発および保守しています。UNSPSCは、商品を分類して識別するスキーマです。これは、購入側カタログと販売側カタログで使用され、支出分析（費用分析）では標準の勘定科目コードとして使用されています。4つのレベルの分類システムがあり、エンド・ユーザーが必要に応じて5番目のレベルを追加できる規定があります。Knowledge Studio には、ユーザーが使用できる UNSPSC 分類タイプのいくつかのバージョンが付属しています。

最新のバージョンを取得する手順は、次のとおりです。

1. UNSPSC の Web サイトを参照します。

<http://www.eccma.org/resources/UNSPSCDownloads.php>

2. 「ECCMA UNSPSC Current file」の横にあるダウンロード・リンクをクリックします。たとえば、「a135UNSPSC.zip - 912 KB」などです。

連絡先情報フォームを含む新しいブラウザ・ウィンドウが表示されます。

3. 連絡先情報を入力し、電子メールで確認コードを取得します。
4. 電子メールで受信した確認番号を入力し、「Verify」をクリックします。
5. 「Click here」リンクをクリックしてダウンロードを開始し、ファイルを保存します。
6. ダウンロードした zip ファイルから Excel ファイルを抽出します。
7. Excel ファイルを開きます。
8. 「BTI」列を削除します。
9. ファイルを .csv ファイルとして保存します。

これで、このファイルを使用して UNSPSC スキーマ分類タイプを作成できます。詳細は、5-3 ページの「分類タイプ」を参照してください。

次に、階層構造である UNSPSC 分類スキーマのサンプルを示します。



この分類システムとそれをサポートする組織の詳細は、UNSSPC の Web サイトを参照してください。

<http://www.unspsc.org>

## eCl@ss

ドイツで開発された eCl@ss は、仕入先とその顧客間での情報交換用の標準分類タイプとして使用されています。eCl@ss の特徴は、4つのレベルの階層分類システムにあります。eCl@ss は、業界の購買担当に市場構造をマップし、開発、計画および保守を担当するエンジニアをサポートします。これは、商品进行分类して識別するスキーマです。これは、販売側カタログと購入側カタログで使用され、支出分析（費用分析）では標準の勘定科目コードとして使用されています。Knowledge Studio には、eCl@ss の最新バージョンが付属しています。この分類システムとそれをサポートする組織の詳細は、eCl@ss の Web サイトを参照してください。

<http://www.eclass-online.com/>

## ユーザー定義

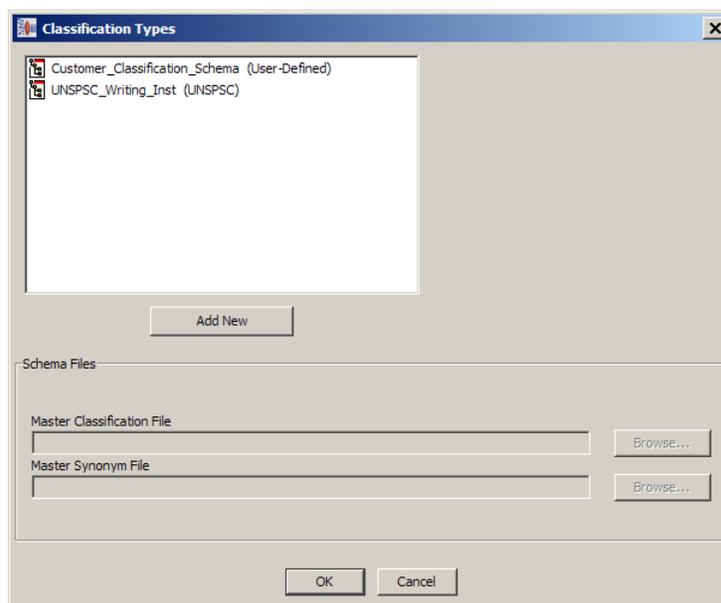
Enterprise DQ for Product には、カスタム（ユーザー定義）分類タイプを作成できるファイル書式が用意されています。

## 分類タイプ

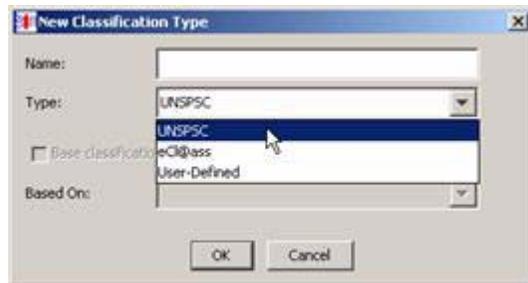
「Classify」タブをアクティブにしてこの機能の使用を開始するには、データ・レンズに対してデフォルトの分類タイプを作成する必要があります。

## 分類タイプの作成

1. 「Data Lens」メニューから「Classification Types...」を選択します。



2. 「Add New」ボタンをクリックします。



3. 分類タイプの名前を入力します。この名前は、選択肢として「**Classification Types**」リストに追加されます。
4. 使用するスキーマのタイプをリストから選択し、新しい分類タイプを作成します。
5. 作成した分類タイプがすでにあり、同じ分類の新しいバージョンでそのナレッジを再利用する場合は、「**Base classification on other classifications**」チェック・ボックスを選択し、次に「**Based On:**」リストから適切な分類タイプを選択します。

---

**注意：** 他の分類タイプがない場合、このチェック・ボックスはアクティブになりません。

---

6. 「OK」をクリックします。  
「**Classification Type**」ダイアログに戻ります。
7. 「**Master Classification File**」を選択し、該当する場合は「**Browse**」ボタンを使用して「**Master Synonym File**」を選択します。

ユーザーは独自のバージョンを使用できますが、各データ・レンズには、e@Class、UNSPSC およびユーザー定義のスキーマ・タイプの一連のサンプルが含まれています。

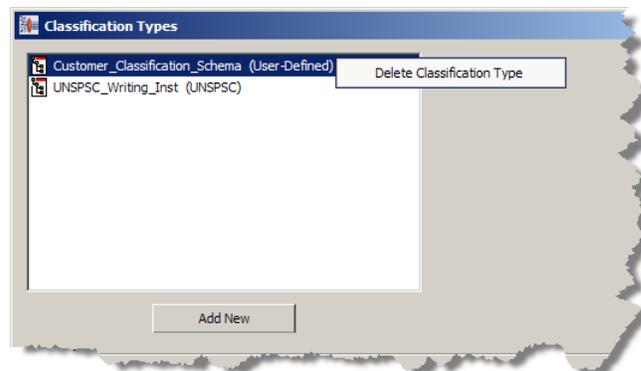
これで、Knowledge Studio の「classification」タブがアクティブになります。

選択した分類タイプは、ツールバーの分類タイプのリストに表示されます。このリストから分類タイプを選択すると、選択したタイプに関連付けられている分類ルールが保守用の「**Classify**」タブに表示されます。「**Phrases**」タブの「**Define Phrases**」サブタブでのフレーズと用語の保守によって、すべての分類タイプが更新されます。プロジェクトを保存すると、分類タイプに加えられたすべての変更が保存されます。

## 分類タイプの削除

分類タイプは、必要に応じて削除できます。

1. 最新バージョンの最新データ・レンズをチェックインしたことを確認します。
2. 「**Data Lens**」メニューから「**Classification Types...**」を選択します。
3. 削除する分類タイプを選択して右クリックします。



4. 「Delete Classification Type」をクリックします。  
削除確認ダイアログが表示されます。
5. 選択した分類タイプを削除する場合は「OK」をクリックし、削除しない場合は「Cancel」をクリックします。
6. 「OK」をクリックします。

データ・レンズ内の分類タイプのみを削除する場合は、「Classify」タブが非アクティブにレンダリングされます。

## 「Classify」タブ

「Classify」タブおよび関連するサブタブには、データの分類に必要なすべての機能が用意されています。

---

**注意：** 分類の親子関係を保守するには、「Classification Context」データ・レンズ・オプションを設定します。詳細は、8-2 ページの「[データ・レンズ・オプションの設定](#)」を参照してください。

---

## 「Classify from Data」 サブタブ

「Classify from Data」 サブタブは、特定のデータと照合してコンテンツを分類するための主要なメカニズムです。

The screenshot displays the 'Classify from Data' interface with three main panes:

- Classification Type Pane:** A tree view on the left showing a hierarchy of categories under 'UNSPSC\_Writing\_Inst'. The selected category is '4121702 - Rollerball pens'.
- Sample Data Table:** A table on the right showing classification results for various items. The first row is highlighted in yellow.
 

#	CC	Category	Item Definition	100 lines of 100 (0 % classified)
1		Mechanical_Pencils	Mechanical_Pencil_SPH Emerald Barrel	
2		Mechanical_Pencils	Pencil_SPH Green Barrel Automatic	
3		Pens	Balpoint Pen Medium Point Black Barrel Black Ink	
4		Pen_Refills	Balpoint Pen Refill Fine Point Blue Ink 2 / Pack	
5		Pen_Refills	Balpoint Pen Refill Med Pt Black Ink 2 / Pk	
6		Pen_Refills	Balpoint Pen Refill Med, Point black, ink 2 / Pack	
7		Pen_Refills	Balpoint Pen Refill Point Mdm Black Ink	
8		Pen_Refills	Balpoint Pen Refill min point, Thin ink Black	
9		China_Markers	China Ink Green Ink 12 / Box	
- Graphical Rule Pane:** A central pane showing a graphical rule tree for 'Mechanical\_Pencils'. The root node is '[a\_barrel\_color]', which branches into '[i\_size]' and '[i\_color]'. '[i\_size]' further branches into '[i\_pencil\_type]' and '[i\_pencil]'. '[i\_pencil\_type]' branches into '[i\_mechanical]' and '[i\_ballpoint]'. '[i\_mechanical]' branches into '[i\_rollerball]', '[i\_ballpoint]', '[i\_ballpoint]', '[i\_ballpoint]', and '[i\_ballpoint]'. The final nodes are 'Mechanical', 'Pencil', 'Ballpoint', 'Ballpoint', and 'Ballpoint'.

### 「Classification Type」 ペイン

「Classification Type」 ペインには、選択した分類スキーマに含まれるカテゴリが表示されます。色付きアイコン、カテゴリのコード番号およびカテゴリの名前は、ツリー・タイプのスキーマの各カテゴリを表します。ツリーの最上位カテゴリには、分類タイプの名前（ここでは「UNSPSC\_Writing\_Inst」）が表示されます。

「Masks」の使用方法については、12-2 ページの「高度な分類ルールの機能」を参照してください。

### 「Sample Data」 ペイン

「Sample Data」 ペインには、選択したサンプル・データ・ファイルのサンプル・データの各行が表示されます。表の各列は、左から右に次の内容を示しています。

#### 行番号 (#)

データ行に割り当てられている一意の番号。

#### 分類件数 (CC)

データ行が分類されていない場合、この列はゼロです。行が1つのカテゴリに分類されている場合、この列には「1」が表示されます。行が2つの異なるカテゴリに分類されている場合、この列には「2」が表示され以下同様です。データ行が複数のカテゴリに分類されている場合、この列はピンク色で色付けされます。

#### Category

行が分類されている場合は、この行のカテゴリ番号が表示されます。カテゴリ識別子をダブルクリックすると、「Classification Type」 ペインの特定のカテゴリに移動します。

#### Item Definition

品目定義がトリガーされている行に対する品目定義。すべてのレコードに対して品目定義が定義されていない場合があります。この場合、この列のデータ行には品目定義がありません。

#### 分類済の行の数 / 割合

製品の摘要。

#### Standardized Text

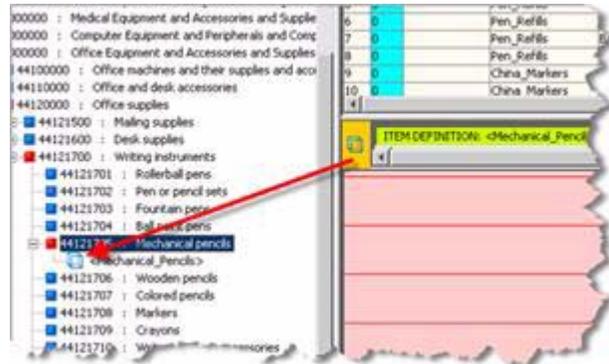
元のデータ行に対する標準化書式。

データが含まれている各列は、そのタイトルをクリックすることで表を昇順と降順の両方でソートできます。列ヘッダーを1回クリックすると、表がソートされ、選択した列の品目別に昇順のアルファベット順で表示されます。同じ列ヘッダーを2回クリックすると、表が再度ソートされ、降順のアルファベット順で表示されます。

## 「Item Definition」 セクション

このセクションのフィールドには、「Sample Data」ペインで選択した行の品目定義が表示されます。このフィールドの任意の場所をダブルクリックすると、「Phrases」タブの「Define Items」サブタブが自動的に選択され、選択した品目定義を変更できます。

品目定義を使用して分類するには、「Item Definition」アイコンを「Classification Type」ペインのカテゴリにドラッグします。



## 「Graphical Rule」 ペイン

「Graphical Rule」ペインには、「Sample Data」ペインで選択した行のフレーズ構造が表示され、分類処理の全コンテキストおよび結果を表示できます。

このペインは、分類変換を作成するためにデータを分類システムにリンクする際のソースとして使用されます。データをリンクするには、ルールまたは品目定義を「Classification Type」ペインのカテゴリ上にドラッグ・アンド・ドロップして分類ルールを作成します。

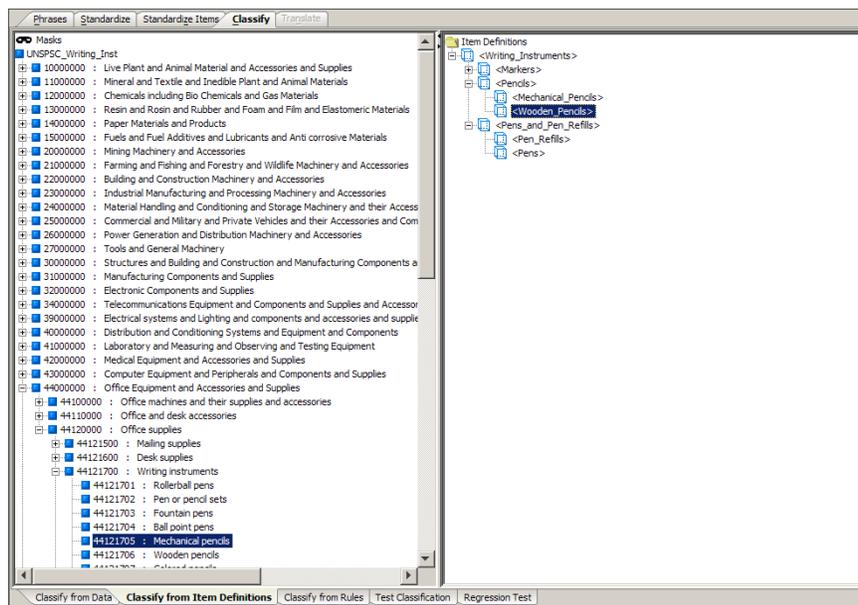
---

**注意：** 品目定義を分類すると、分類した品目の分類階層のカテゴリ・ノード・アイコンが青色から赤色に変更されます。

---

## 「Classify from Item Definitions」 サブタブ

「Classification by Item Definitions」 サブタブは、外部の分類法またはスキーマと照合してコンテンツを分類する主要なメカニズムで、品目定義は Enterprise DQ for Product のコンポーネントとして重要性が増しています。品目定義に、データ行の分類に必要なすべての情報が含まれている場合は、データの分類にルールを使用する必要はありません。ただし、品目定義構造が分類スキーマより低い粒度の場合、分類の実行には用語およびフレーズのルールも必要になる場合があります。



右側のペインの「Item Definitions」を「Classification Type」ペインのカテゴリに分類するには、「Item Definition」アイコンを左側の適切なカテゴリにドラッグします。

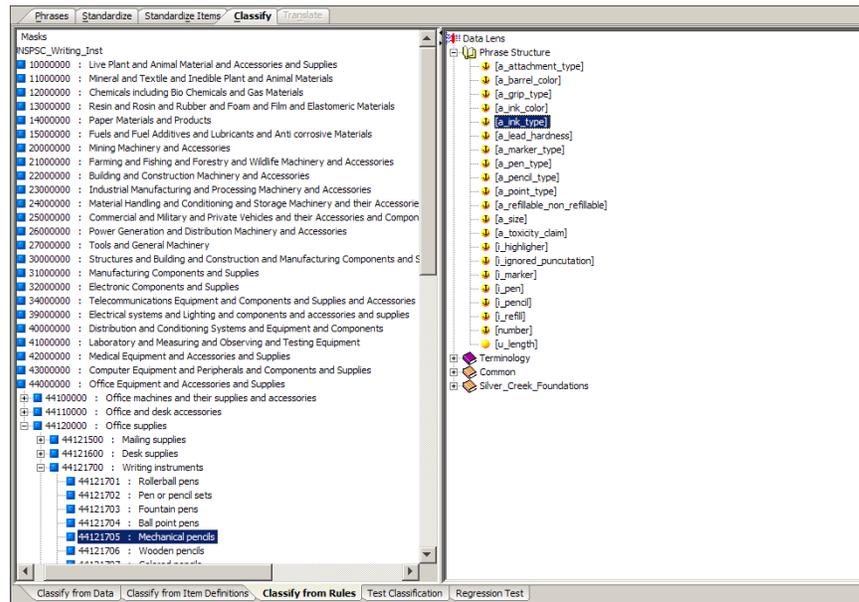
品目定義は、同じ方法を使用して分類階層内の必要な数のカテゴリに分類できます。これは、複数分類と呼ばれます。



品目定義または品目定義とルールの分類には、複数分類を使用できます。複数の分類ノードにルールのみをドラッグした場合、複数分類は機能しません。ルールは分類階層の2つの異なるノードに分類できますが、DSA や DGS プロジェクトなどに出力されるのは最初の分類のみです。

## 「Classify from Rules」サブタブ

「Classify from Rules」サブタブでは、フレーズおよび用語をカテゴリにリンクできます。



分類ルールを作成するには、右側のペインでフレーズを選択し、左側の「Classification Type」ペインのカテゴリ上にドラッグ・アンド・ドロップします。「Classify from Data」サブタブには、分類したデータ行が表示されますが、このサブタブにはデータ・レンズのフレーズが表示されます。

データを分類するときは、できるかぎり慎重に選択します。UNSPSC 分類タイプは4つのレベルのシステムであるため、品目定義は、可能なかぎりを4つのレベルに分類することをお勧めします。ただし、分類システムは完全とはかぎらないため、品目を一般的なカテゴリに分類することが必要になる場合があります。

## 「Test Classification」 サブタブ

「Test Classification」 サブタブでは、サンプル・データの分類をレビューして結果を検証できます。

#	QID	Lien	Initial Text	Standardized Text
1	100	42	Mechanical Pencil, 3MM Emerald Barrel	Mechanical, Pencil, Office, Emerald Barrel
2	1000	79	Pencil, 3MM Green Barrel Automatic	Automatic, Pencil, Office, Green Barrel
3	98	53	Ballpoint Pen Medium Point Black Barrel Black Ink	Pen, Ballpoint, Black Ink, Black Barrel, Medium Point
4	88	51	Ballpoint Pen Refill Fine Point Blue Ink 2 / Pack	Ballpoint, Blue Ink, Fine Point, Pen, 2 per package
5	88	48	Ballpoint Pen Refill Med Pt Black Ink 2 / PK	Ballpoint, Black Ink, Med Pt, Pen, 2 per package
6	88	53	Ballpoint Pen Refill Meds, Point black, PK 2 / Pack	Ballpoint, black ink, Meds, Point, Pen, 2 per package
7	88	36	Ballpoint Pen Refill Point Mdm Black Ink	Ballpoint, Black Ink, Point Mdm, Pen
8	100	52	Ballpoint Pen Refill mdm point - 7mm ink Black	Ballpoint, ink Black, 0.7 millimeter, mdm point, Pen
9	100	38	China Mkr Green Ink 12 / Box	China, Mkr, Green Ink, 12 per package
10	100	23	China Markers red ink	China, Markers, red ink
11	100	25	China Markers White ink	China, Markers, White ink
12	100	22	China Mkr yellow ink	China, Mkr, yellow ink
13	100	40	Chisel Point Highlighter Turquoise Ink	Highlighter, Chisel Point, Turquoise ink
14	88	64	Comfort Grip Retractable Ballpoint Pen medium Point Blue ink	Pen, Retractable Ballpoint, Blue ink, Comfort Grip, medium Point
15	88	63	Comfort Grip Retractable Ballpoint Pen Medium Point Red Ink	Pen, Retractable Ballpoint, Red Ink, Comfort Grip, Medium Point
16	75	40	Counter top pen refill Medium Point black ink	black Ink, Medium Point, Counter top, pen
17	100	60	Dry-Erase Chisel Point Markers Black Ink 2 / Set	Markers, Dry - Erase, Black Ink, Chisel Point, 2 per package
18	100	57	Dry-Erase Chisel Tip Markers Assorted 4 / Set	Markers, Dry - Erase, Assorted, Chisel Tip, 4 per package
19	75	37	Security Pen Refills Medium Point Blue Ink	Blue Ink, Medium Point, Security, Pen
20	75	36	Counter Pen Refills Medium Point Blue Ink	Blue Ink, Medium Point, Counter, Pen
21	100	43	Dry-Erase Marker Bullet Point Blue Ink	Marker, Dry - Erase, Blue Ink, Bullet Point
22	100	41	Dry-Erase Marker Fine Point Blue Ink	Marker, Dry - Erase, Blue Ink, Fine Point
23	100	39	Dry Erase Markers Chisel Tip Red Ink	Markers, Dry Erase, Red Ink, Chisel Tip
24	88	55	Retractable Ballpoint Pen Non Refillable Black Ink	Pen, Retractable Ballpoint, Black Ink, Non - Refillable

Source Text (editable - press Enter when done)  
 Ballpoint Pen Medium Point Black Barrel Black Ink

Standardized Text    Attribute Separator    Append Unattributed Text    Append Unparsed Text

Pen, Ballpoint, Black Ink, Black Barrel, Medium Point

Item Attributes

Pens: Product\_Name=Pen, Type=Ballpoint, Ink\_Color=Black Ink, Size=nu8, Ink\_Type=nu8, Barrel\_Color=Black Barrel, Grip\_Type=nu8, Point\_Type=Medium Point.

Standardize Attributes    Order Attributes    Match Weights    Test Attributes    Test Item Standardization    Regression Test

## 「Sample Data」 ペイン

この表には、元のデータと分類された後の同じデータが表示されます。各列は、「Classify from Data」サブタブのサンプル・データ表（5-6 ページの「「Classify from Data」サブタブ」を参照）と同じですが、右側の最後の列には、標準化されたデータが表示されます。

## ソース・フィールド

このフィールドには、元のデータが表示されます。このフィールドは編集可能で、[Enter] キーを押すと、データ・レンズ上で即時に結果をレビューできます。

## 「Standardized Text」 フィールド

このフィールドには元のデータの標準化バージョンが表示されますが、編集はできません。

## 「Classification」 ペイン

「Classification Type」 ペインのサブセットには、選択した明細品目の分類先が表示されます。

## 複数分類のテスト

分類件数が複数ある場合は、ツールバーの前へ / 次へ矢印がアクティブになります。これらの矢印を使用すると、「Classification」ペインに表示する分類を変更できます。「Sample Data」ペインのカテゴリ名は変更されませんが、次の分類を表示するテスト・ペインでは変更されます。「Sample Data」ペインの「CC」列には、複数分類の件数が表示されます。

## 「Regression Test」 サブタブ

回帰テストはデータ標準化の重要な部分で、データの分類が予想どおりであることを確認できます。

タブがアクティブでない場合は、「**Regression Testing Active**」データ・レンズ・オプションを設定します。詳細は、2-28 ページの「**回帰テストがアクティブであることの確認**」を参照してください。

#	CC	Regression	Standardize	Standardize Items	Classify	Translate	ID	Standardized Data
45	0						W153	Highlighter Chisel Point Fluorescent Orange
44	0						W152	Highlighter Chisel Point Fluorescent Green
43	0						W151	Highlighter Chisel Point Blue
42	0						W145	Highlighter Chisel Point Assorted Colors 5 per package
41	0						W144	Highlighter Chisel Point 4 per package
40	0						W143	Roller Pen Refill Fine Point Violet Ink
39	0						W142	Gel Ink Roller Pen Refill Fine Point Orange Ink
38	0						W141	Automatic Pencil 0.5 millimeter Kryptonite Green Barrel
37	0						W140	Automatic Pencil with Eraser / Metal Clip 0.5 millimeter Blue Cap
36	0						W139	Automatic Pencil with Eraser / Metal Clip 0.5 millimeter Black Cap
35	0						W138	Gel Ink Roller Ball Pen Refill 0.7 millimeter Black Ink 2 per package
34	0						W137	Gel Grip Rollerball Pen Refill 1 millimeter Red
33	0						W136	Fine Tip Permanent Marker 1 millimeter Lime Ink
32	0						W135	Highlighter Chisel Tip Yellow
31	0						W134	Highlighter Chisel Tip Pink
30	0						W133	Highlighter Chisel Tip Orange
29	0						W132	Highlighter Chisel Tip Fluorescent Yellow
28	0						W131	Highlighter Chisel Tip Fluorescent Purple
27	0						W130	Felt Tip Pen Extra - Fine Point Translucent black Barrel Red Ink
26	0						W129	Felt Tip Pen Extra - Fine Point Translucent blue Barrel Purple Ink
25	0						W128	Erasable Ink Ballpoint Pen Med Point Blue Barrel / Blue Ink
24	0						W127	Retractable Ballpoint Pen Non - Refillable Black Ink
23	0						W126	Dry Erase Markers Chisel Tip Red Ink
22	0						W125	Dry - Erase Marker Fine Point Blue Ink
21	0						W124	Dry - Erase Marker Bullet Point Blue Ink
20	0						W123	Counter Pen Refills Medium Point Blue Ink
19	0						W122	Security Pen Refills Medium Point Blue Ink
18	0						W121	Dry - Erase Chisel Tip Markers Assorted 4 per package
17	0						W120	Dry - Erase Chisel Point Markers Black Ink 2 per package
16	0						W119	Countertop pen refill Medium Point black Ink
15	0						W118	Comfort Grip Retractable Ballpoint Pen Medium Point Red Ink
14	0						W117	Comfort Grip Retractable Ballpoint Pen Medium Point Blue Ink
13	0						W116	Chisel Point Highlighter Turquoise ink
12	0						W115	China Mkr yellow ink
11	0						W114	China Markers White ink
10	0						W113	China Markers red ink
9	0						W112	China Mkr Green Ink 12 per package
8	0						W111	Ballpoint Pen Refill mdm point 0.7 millimeter Ink Black

「**Sample Data**」 ペインはこのサブタブで構成され、「**Test Classification**」 サブタブ上のペインと同じです。

サンプル・データ表にデータが表示されない場合は、サンプル・データが初期化されていないため、回帰ベースは存在しません。回帰ベースの初期化の詳細は、5-11 ページの「**回帰ベースの作成と更新**」を参照してください。

## 回帰ベースの作成と更新

回帰ベースの作成におけるベスト・プラクティスは、サンプル・データを1つのファイルに結合することです。詳細は、8-12 ページの「**サンプル・ファイルの結合**」を参照してください。

ファイルの結合によってデータが削除されることはなく、選択したサンプル・ファイルが新規の大規模なファイルに単純に結合されます。

次に、回帰ベースのサンプル・データ・ファイルに対して1つの変更を加え、回帰セットを確認して適切に更新します。複数の変更を加えると回帰の結果が読みにくくなり、エラーを見落としたり修正が困難になる可能性があります。

回帰ベースを作成するには、「**File**」メニューの「**Create New Regression Base**」オプションを選択し、回帰テストに使用するサンプル・データ・ファイルを選択します。これによって、回帰ベースが初期化され、「**After**」ペインに結果が表示されます。

回帰ベースは、「**File**」メニューの「**Update Regression Base**」オプションを使用して、レビュー済および受入済のテキスト行（前述の赤色チェック・マーク列）で更新できます。

---

**注意：** サンプル・データをレビューまたは受け入れた場合のみ、回帰ベースを初期化または更新してください。

---



---

---

## データの翻訳

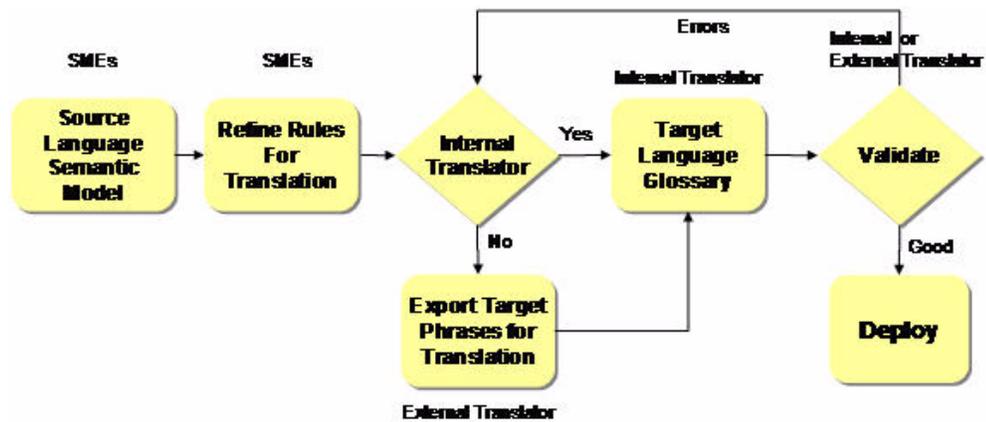
Enterprise DQ for Product では、データ・レンズ標準化の結果を利用して、製品データをバッチまたはリアルタイムで翻訳できます。また、1つのソース言語から一連の標準化属性を取得して、対応するターゲット言語翻訳を翻訳用語集に格納できます。この用語集ファイル全体は、Knowledge Studio を使用して開発でき、エクスポート・ユーティリティを使用してエクスポートし、翻訳のために翻訳者に送信できます。完成した翻訳は、Knowledge Studio にインポートして戻すことができます。

翻訳用語集は、次の翻訳を生成するためにリアルタイムまたはバッチで使用されます。

- ソース・データ・レンズに基づいた、ターゲット言語（1つまたは複数）への属性の構造化コンテンツの翻訳。
- 翻訳済属性に基づいた、ターゲット言語（1つまたは複数）への摘要の翻訳。

標準化と同様に、データ・レンズおよびユースケースに基づいて解析する必要があるのは、キー属性のみです。翻訳では、データ・レンズの作成と認識、およびプロジェクトの標準化フェーズで実行したすべての作業が利用されます。この差異がターゲット言語の要件です。

## 翻訳プロセス



### データ・レンズ標準化の品質

翻訳で重要な点は、翻訳プロセスに備えて、標準化属性を使用した完全なデータ・レンズを設定することです。翻訳は、現在選択されている単位換算および標準化に基づいています。翻訳の前に、影響するデータ・レンズ内のすべての標準化の品質を慎重にレビューして確認する必要があります。

### 翻訳に備えたデータの準備

一部のテキストは翻訳が不要です。たとえば、コード、数字および固有名詞などです。「Do Not Translate」ルール属性では、コンテキスト内の各属性に対するセマンティック・モデル解釈を利用します。これには、数字、コード、固有名詞などの処理ルールが含まれます。この属性はデータ・レンズを介してフローし、翻訳用語集へのエントリが不要なため、翻訳する必要があるルールの数が減少します。

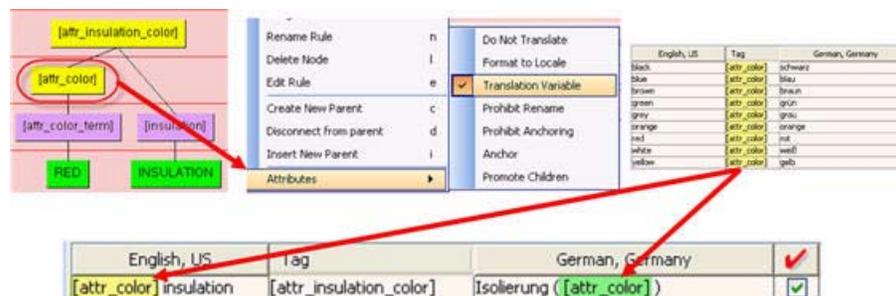


反対に、他のテキストではターゲットのローカル構文に対応する書式設定が必要です。たとえば、数字や通貨などです。データ・レンズでこのテキストを識別すると、単位の翻訳コストが削減されます。つまり、単位が翻訳されるのは1回のみで、Knowledge Studio はターゲット言語の要件に基づいて数字を自動的に書式設定します。



English, US	Tag	German, Germany	
[integer] mm squared	[attr_conductor_size]	[integer] mm2	✓
[integer;1] mm squared , [integer;2] Conductor , [attr_cable_conductors]		[integer;1] mm2, [integer;2] Leiter,	✓

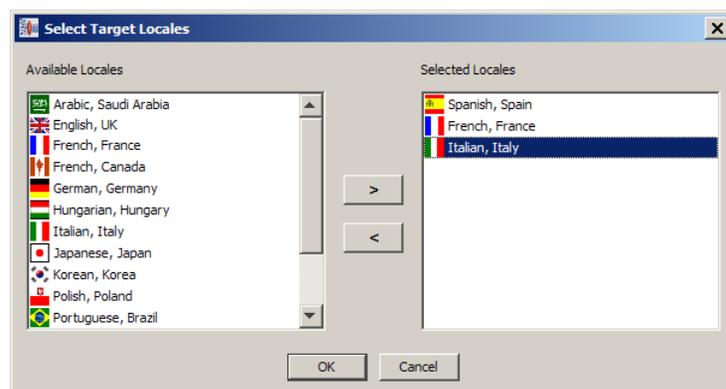
さらに、翻訳変数として使用されるテキストがあります。たとえば、色や材料などです。これによって、異なる属性の順序が必要なフレーズの翻訳コストが削減されます。次に、個々の色を1回のみ翻訳する [attr\_color] フレーズの例を示します。[attr\_color] フレーズには、特別な翻訳属性の「Translation Variable」が指定されています。これによって、[attr\_color] の用語を使用する上位レベル属性では個々の色の翻訳を再利用することが Knowledge Studio に通知されます。



## 翻訳ターゲット

データ・レンズを完全に標準化した後は、翻訳ターゲットを識別して「Translation」タブをアクティブにする必要があります。

「Data Lens」メニューから「Translation Targets」をクリックします。



「Select Target Locales」ダイアログでは、「Available Locales」リストからデータ・レンズの翻訳言語を1つ以上選択できます。

左側のリストには使用可能なすべてのロケールが、右側のリストにはアクティブなすべてのロケールが表示されます。2つのリスト間にある矢印ボタンを使用してリスト間でロケールを移動し、選択を完了します。

選択済ロケールがツールバー上の「Translation Targets」リストに移入されます。

1つのデータ・レンズにある定義済および標準化したフレーズと用語を再利用して任意の数の翻訳結果を定義できるように、必要な数の翻訳ターゲットを選択できます。

翻訳ターゲットを選択した後は、ツールバー上の「Translation Targets」リストを使用して、編集する任意のターゲットを選択でき、「Initial Translation」ボタンの外観も選択した言語のアイコンに変更されます。「Phrases」タブでのフレーズと用語の保守、および「Standardize」タブでの標準化によって、すべての翻訳ターゲットが更新されます。プロジェクトを保存すると、翻訳ターゲットに加えられたすべての変更が保存されます。

## 翻訳スマート用語集

ツールバー上の「Initial Translation」ボタンをクリックすると、翻訳スマート用語集をインポートしてそのナレッジを適用し、翻訳プロセスを高速化できます。リストから適切な翻訳用語集を選択します。



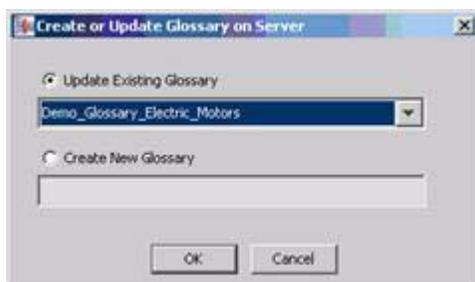
Oracle DataLens Server の翻訳用語集は次のディレクトリにあります。

```
...¥datalens¥server¥data¥shared¥common¥glossary
```

## Oracle DataLens Server の翻訳スマート用語集の作成 / 更新

Oracle DataLens Server の翻訳スマート用語集は、既存の内容を更新することも、データ・レンズの完成済翻訳に基づいて新規に作成することもできます。

「File」メニューから「Create/Update Glossary」をクリックします。



適切な処理を選択します。更新するには、リストから既存の用語集を選択する必要があります。新規の翻訳用語集を作成するには、新規用語集のわかりやすい名前を入力します。

---

**注意：** エラーを回避するには、「Export」チェック・ボックスを使用して1つ以上の行を選択しておく必要があります。

---

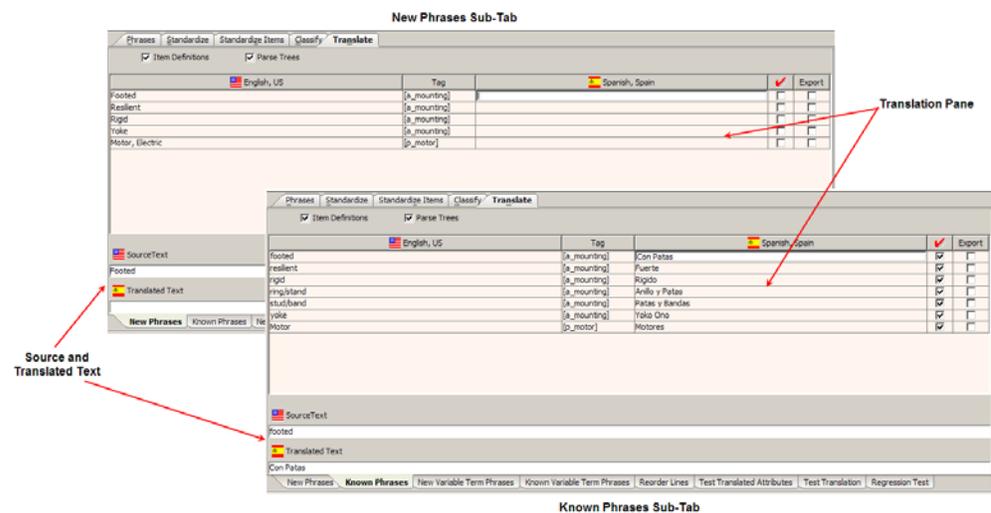
## 「Translation」タブ

「Translation」タブでは、翻訳データを作成できます。翻訳用語集の翻訳を生成するには、3つの方法があります。

- 「Translation」タブのサブタブに翻訳済テキストを手動で入力します。
- 翻訳のすべてのデータをエクスポートし、次にすべての翻訳済データをデータ・レンズにインポートします。
- エクスポート / インポートに対して、翻訳済データの手動入力と特定データの選択を組み合わせます。

## 「New Phrases」サブタブおよび「Known Phrases」サブタブ

翻訳は、翻訳テキストを「New Phrases」サブタブに直接入力して作成できます。翻訳済フレーズを入力するときは、赤色のチェック・ボックスを選択してこの翻訳を文法に含めることを指定し、完全な翻訳済フレーズが「Known Phrases」サブタブに追加されるように適用します。



これらのサブタブには、同じ情報および機能が同様の方法で含まれています。

## 「Translation」ペイン

それぞれの「Translation」ペインの左側にはソース言語テキストが、右側には翻訳テキストが表示されます。

このペインに表示されるデータは、「Item Definitions」および「Parse Tree」チェック・ボックスを次のように切り替えることで変更できます。

### Item Definitions

リストに品目定義およびフレーズが表示されます。

### Parse Tree

選択したテキストの品目定義階層が検索されます。

データ・レンズを保存すると、フレーズ翻訳の変更がそのデータ・レンズ用ナレッジ・ベースのロケール・ディレクトリにある翻訳用語集に書き込まれます。

### 赤色のチェック・マーク

赤色のチェック・マークのチェック・ボックスを使用して、「New Phrases」サブタブでレビューした翻訳済テキストの行を指定すると、その行が「Known Phrases」サブタブで自動的に選択されます。

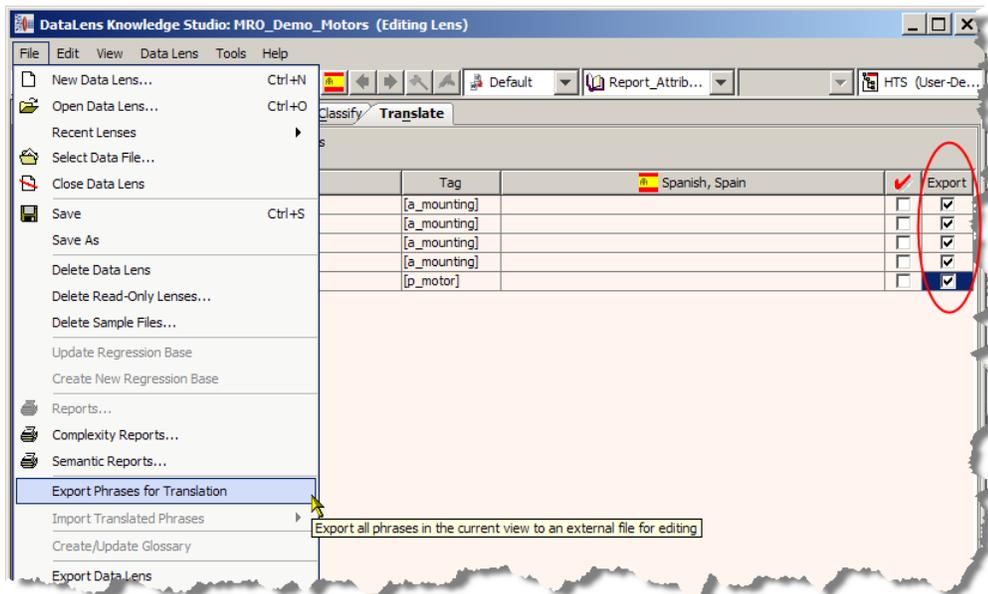
## Export

いずれかのサブタブで「Export」チェック・ボックスを使用すると、外部翻訳のリストにエクスポートするテキスト行を指定できます。

外部翻訳を実行するには、翻訳が必要なフレーズのみを単にエクスポートします。このプロセスでは、必要な属性の翻訳は1回のみのため、時間と費用を節約できます。データ・レンズでは、すでに翻訳されている属性を再利用できます。翻訳した属性は、Knowledge Studio にインポートできます。

## 翻訳のためのテキストのエクスポート

「New Phrases」サブタブから翻訳するフレーズをエクスポートするには、エクスポートする各フレーズの「Export」チェック・ボックスを選択します。「File」メニューから「Export Phrases for Translation」を選択します。



これによって、...¥data lens name¥Application Data¥DataLens¥export エクスポート・ディレクトリにデータ・レンズ名と .trn 拡張子が指定されたファイルが生成されます。

エクスポートされる翻訳ファイル形式は Unicode のタブ区切り形式であるため、Unicode と互換性のあるプログラムで編集する必要があります。Microsoft Excel は、このファイルを Unicode テキストで保存できるプログラムの 1 例です。

このファイルには、次の内容をタブ区切り形式で設定します。

ソース	フレーズ・タグ	ターゲット
footed	[a_mounting]	Con Patas
rigid	[a_mounting]	Rigido
ring/stand	[a_mounting]	Anillo y Patas

## 翻訳のためのテキストのインポート

翻訳されたフレーズは、データ・レンズにインポートできます。

この機能は「**Known Phrases**」サブタブでのみアクティブであるため、このサブタブを選択する必要があります。インポートの選択肢には、次の2つがあります。

### Import Current Translated Phrases

赤色のチェック・マークのチェック・ボックスを使用して選択した翻訳済フレーズのみがインポートされます。

### Import All Translated Phrases

インポート・ファイルに含まれているすべての翻訳済フレーズがインポートされます。

「**File**」メニューから「**Import Translated Phrases**」を選択し、適切なインポート選択肢を選択します。デフォルトの場所はユーザーのエクスポート・ディレクトリですが、翻訳済テキスト・ファイルの正確な場所を選択できます。

インポートすると、確認ボックスが選択された状態の「**Known Phrases**」タブに翻訳が表示されるため、正確性について最終的なレビューを完了できます。

## 「Source Text」フィールドおよび「Translated Text」フィールド

これらのフィールドには、「**Translation**」ペインで選択したデータ・ソースとそのソースの翻訳テキストが表示されます。「**Source Text**」フィールドは、編集できません。「**Translated Text**」フィールドは編集可能で、すべての変更は「**Translation**」ペインに反映されます。

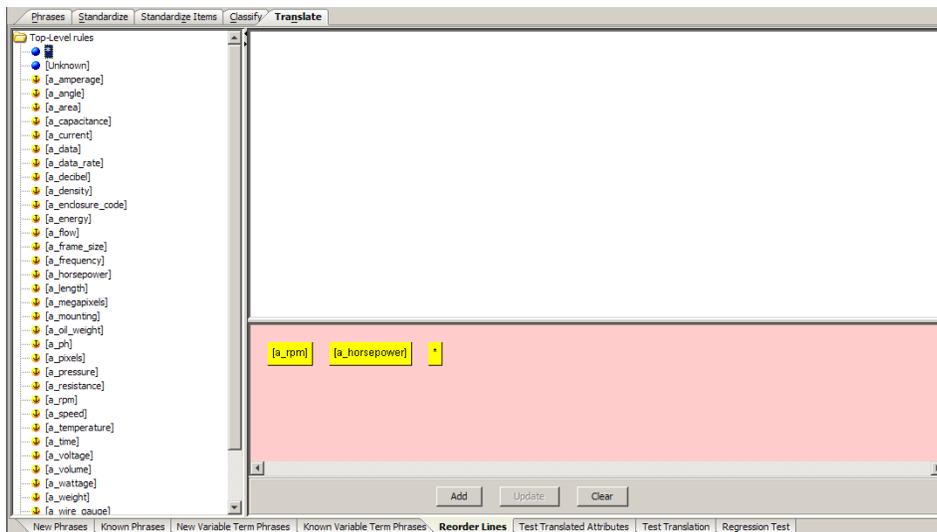
フレーズをエクスポートし、言語翻訳者に提供します。

## 「New Variable Term Phrases」サブタブおよび「Known Variable Term Phrases」サブタブ

これらのサブタブは「**Known Phrases**」サブタブのように動作しますが、表示されるデータは異なります。「**New Variable Term Phrases**」サブタブには新たに翻訳されたすべての変数が表示されますが、「**Known Variable Term Phrases**」サブタブにはデータ・レンズに対しては既知の翻訳済変数すべてが表示されます。

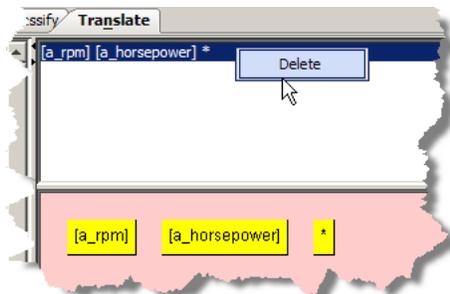
## 「Reorder」サブタブ

言語翻訳者または知識のある他のユーザーは、「Reorder Lines」サブタブを使用して最終的な属性の順序を完了し、翻訳済フレーズの順序がターゲット言語で文法的に正しいことを確認できます。



左側のタスク・ペインからフレーズを選択して「Graphical Rule」ペインに表示すると、ターゲット翻訳ロケール用に適切に順序変更したフレーズを形成できます。順序変更済フレーズを完了した後は、「Add」をクリックしてデータ・レンズに追加します。

追加した順序変更済フレーズは、「Graphical Rule」ペインで消去することも、右側のタスク・ペインで右クリックし、「Delete」を選択して削除することもできます。



次の項では、翻訳テストのサブタブを使用して順序変更の修正をテストする方法について説明します。

## 「Test Translation」 サブタブ

最終的な品質保証の確認は、「Test Translated Attributes」、「Test Item Translation」および「Test Global Translation」の各サブタブを使用して実行できます。未翻訳のフレーズは青色に色付けされ、翻訳済フレーズは白色に色付けされます（色の強調表示はありません）。

デフォルトでは、アクティブと非アクティブの両方の品目定義のデータが表示されます。アクティブな品目定義のデータ（生成結果データ）のみを表示するには、「View」メニューから「View as Production」を選択します。非アクティブな品目定義データの表示の詳細は、2-18ページの「アクティブまたは非アクティブな品目定義」を参照してください。

## 「Test Translated Attributes」 サブタブ

「Test Translated Attributes」サブタブでは、品目定義の翻訳済フレーズの属性をレビューして結果を検証できます。このサブタブは、「Standardize Items」タブの「Test Attributes」サブタブに類似しています。

The screenshot displays the 'Test Translated Attributes' sub-tab. At the top, there are tabs for 'Phrases', 'Standardize', 'Standardize Items', 'Classify', and 'Translate'. The 'Translate' tab is active. Below the tabs, there is a dropdown menu for 'Item Definition' set to '<Electric\_Motors>'. There are two checkboxes: 'Show Value' (checked) and 'Show Text' (unchecked).

The main area contains a table with columns: #, QTY, Text, Item\_Name, Horse\_Power, RPM, Frame\_Size, Voltage, Enclosure\_c..., and Mounting. The table has 5 rows of data. Row 1 is highlighted in blue. Row 2 is highlighted in yellow. Row 3 is highlighted in red. Row 4 is highlighted in blue. Row 5 is highlighted in blue.

Below the table, there is a section for 'Source Text' with a radio button 'Source Text' selected. The text is '1/2HP MTR ODP 1726RPM 115/230V 48 RESILIENT'. Below that is a section for 'Translated Text' with a radio button 'Translated Text' selected. The text is 'Motor, Electric 0.333333 HP 1725 RPM 115 / 230 Volt ODP Resilient'. There are also checkboxes for 'Attribute Separator', 'Append Unattributed Text', and 'Append Unparsed Text'.

At the bottom, there is a 'Standardized Attributes Table' with columns: Name, Value, Text, Numeric Value, and Unit of Measure. The table has 7 rows of data. Row 1 is highlighted in blue. Row 2 is highlighted in red. Row 3 is highlighted in red. Row 4 is highlighted in blue. Row 5 is highlighted in blue. Row 6 is highlighted in blue. Row 7 is highlighted in blue.

At the very bottom, there are several tabs: 'New Phrases', 'Known Phrases', 'New Variable Term Phrases', 'Known Variable Term Phrases', 'Reorder Lines', 'Test Translated Attributes' (active), 'Test Translation', and 'Regression Test'.

## 「Item Definition」 セクション

このリストには、このサブタブを選択する前に選択した品目定義が表示されます。データ・レンジ内のすべての品目定義が選択用にリストされます。

「Show Value」オプションおよび「Show Text」オプションを使用すると、データの表示方法を変更できます。

## サンプル・データ表

この表には、元のデータと標準化された後の同じデータが表示されます。各列は、左から右に次の内容を示しています。

### 行番号 (#)

データ行に割り当てられている一意の番号。

### 品質指標 (QI)

0 から 100 までの数値で、行が標準化された程度を表しています。

### 赤色のチェック・マーク

レビューしてマーク付けしたデータで、該当するデータ行でチェック・ボックスを選択します。

### 元テキスト

データ・レンズで解析された元のデータ。

### 残りの各列

残りの各列は、各品目定義の属性に依存するため、これらの列は変化します。

データが含まれている各列は、そのタイトルをクリックすることで表を昇順と降順の両方でソートできます。列ヘッダーを 1 回クリックすると、表がソートされ、選択した列の品目別に昇順のアルファベット順で表示されます。同じ列ヘッダーを 2 回クリックすると、表が再度ソートされ、降順のアルファベット順で表示されます。

サンプル・データ表内の行のいずれかを選択すると、**標準化属性表**で選択した行の品目定義情報が表示されます。

## ソース・フィールド

このフィールドには、元のデータが表示されます。このフィールドは編集可能で、**[Enter]** キーを押すと、データ・レンズ上で即時に結果をレビューできます。

## 「Translated Text」 セクション

このフィールドには元のテキストの翻訳済バージョンが表示されますが、編集はできません。

「Attribute Separator」 フィールドには、属性間で使用するテキスト・セパレータを入力できます。

「Append Unattributed Text」 チェック・ボックスを選択すると、属性でないすべてのテキストが摘要に挿入されます。

同様に、「Append Unparsed Text」 チェック・ボックスでは、未解析のテキストが摘要に挿入されます。

## 標準化属性表

この表には、選択した品目定義のすべての標準化属性および「Sample Data」 ペインで選択した行が表示されます。このフィールドの表示方法には、「Translated Text」 セクションから選択した属性標準化が反映されます。

## 「Test Item Translation」 サブタブ

「Test Item Translation」サブタブでは、品目定義の翻訳済フレーズをソース言語のサンプル・データと照合してレビューし、結果を検証できます。このサブタブは、「Standardize Items」タブの「Test Item Standardization」サブタブに類似しています。

#	QI	Len	Standardized Text	Translated Text
1	0	16	0.5 MM, BURGLINDY	0.5 MM, BURGLINDY
2	0	23	0.9 MM, GREEN, KRYPTONITE	KRYPTONITE, 0.9 MM, GREEN
3	0	25	0.5 MM, METAL, YES, BLACK	0.5 MM, YES, METAL, BLACK
4	0	24	0.5 MM, METAL, YES, BLUE	0.5 MM, YES, METAL, BLUE
5	0	20	BLACK, MEDIUM, BLACK	Pen: Inactive in Production
6	0	24	BALLPOINT, BLUE, FINE, 2	Pen_refills : Inactive in Production
7	0	32	BALLPOINT, BLACK, MEDIUM, 0.7 MM	Pen_refills : Inactive in Production
8	0	27	BALLPOINT, BLACK, MEDIUM, 2	Pen_refills : Inactive in Production
9	0	27	BALLPOINT, BLACK, MEDIUM, 2	Pen_refills : Inactive in Production
10	0	24	BALLPOINT, BLACK, MEDIUM	Pen_refills : Inactive in Production
11	0	16	BLACK, FINE, YES	Pen: Inactive in Production
12	0	24	BLACK, FINE, RUBBER, YES	Pen: Inactive in Production
13	0	26	BLUE, MEDIUM, COMFORT, YES	Pen: Inactive in Production
14	0	25	RED, MEDIUM, COMFORT, YES	Pen: Inactive in Production
15	0	24	COUNTERTOP, BLUE, MEDIUM	Pen_refills : Inactive in Production
16	0	25	COUNTERTOP, BLACK, MEDIUM	Pen_refills : Inactive in Production
17	0	12	0.7 MM, BLUE	0.7 MM, BLUE
18	0	6	0.5 MM	0.5 MM
19	0	15	0.5 MM, EMERALD	0.5 MM, EMERALD
20	0	26	0.5 MM, TRANSPARENT VIOLET	0.5 MM, TRANSPARENT VIOLET
21	0	14	0.7 MM, A66C6FFF	0.7 MM, A66C6FFF

### サンプル・データ表

この表の複数の列は、「Test Translated Attributes」サブタブ（6-9 ページの「「Test Translated Attributes」サブタブ」を参照）の列と同じで、同じように動作します。異なる列は次のとおりです。

#### 長さ (Len)

元のテキストの文字列の長さを示します。

#### Standardized Text

標準化の元のテキストのソース言語バージョンを示します。

#### Translated Text

標準化の元のテキストの翻訳済バージョンを示します。

#### ソース・フィールド

このフィールドには、ソース言語テキストが表示されます。このフィールドは編集可能で、[Enter] キーを押すと、データ・レンズ上で即時に結果をレビューできます。

#### 「Item Definition」フィールド

このフィールドには選択した品目定義のすべての属性が表示されますが、編集はできません。このフィールドの表示方法には、「Translated Text」セクションから選択した属性標準化が反映されます。

## 「Test Global Translation」 サブタブ

「Test Global Translation」サブタブでは、品目定義の翻訳済フレーズをソース言語のサンプル・データと照合してレビューし、結果を検証できます。このサブタブは、「Standardize」タブの「Test Global Standardization」サブタブに類似しています。

#	QI	Len	English, US	Spanish, Spain
1	D	33	0.5 MM BURGUNDY MECHANICAL PENCIL	0.5 MM BURGUNDY MECHANICAL PENCIL
2	D	40	AUTOMATIC PENCIL 0.5 MM KRYPTONITE GREEN	AUTOMATIC PENCIL 0.5 MM KRYPTONITE GREEN
3	D	46	AUTOMATIC PENCIL WITH YES METAL 0.5 MM BLACK	AUTOMATIC PENCIL WITH YES METAL 0.5 MM BLACK
4	D	45	AUTOMATIC PENCIL WITH YES METAL 0.5 MM BLUE	AUTOMATIC PENCIL WITH YES METAL 0.5 MM BLUE
5	D	32	BALLPOINT PEN MEDIUM BLACK BLACK	BALLPOINT PEN MEDIUM BLACK BLACK
6	D	32	BALLPOINT PEN REFILL FINE BLUE 2	BALLPOINT PEN REFILL FINE BLUE 2
7	D	40	BALLPOINT PEN REFILL MEDIUM 0.7 MM BLACK	BALLPOINT PEN REFILL MEDIUM 0.7 MM BLACK
8	D	35	BALLPOINT PEN REFILL MEDIUM BLACK 2	BALLPOINT PEN REFILL MEDIUM BLACK 2
9	D	35	BALLPOINT PEN REFILL MEDIUM BLACK 2	BALLPOINT PEN REFILL MEDIUM BLACK 2
10	D	33	BALLPOINT PEN REFILL MEDIUM BLACK	BALLPOINT PEN REFILL MEDIUM BLACK
11	D	28	BALLPOINT PEN YES FINE BLACK	BALLPOINT PEN YES FINE BLACK
12	D	35	BALLPOINT PEN RUBBER FINE YES BLACK	BALLPOINT PEN RUBBER FINE YES BLACK
13	D	37	COMFORT YES BALLPOINT PEN MEDIUM BLUE	COMFORT YES BALLPOINT PEN MEDIUM BLUE
14	D	26	COMFORT YES BALLPOINT PEN MEDIUM RED	COMFORT YES BALLPOINT PEN MEDIUM RED
15	D	33	COUNTERTOP PEN REFILL MEDIUM BLUE	COUNTERTOP PEN REFILL MEDIUM BLUE
16	D	34	COUNTERTOP PEN REFILL MEDIUM BLACK	COUNTERTOP PEN REFILL MEDIUM BLACK
17	D	35	MECHANICAL PENCIL 0.7 MM CLEAR BLUE	MECHANICAL PENCIL 0.7 MM CLEAR BLUE
18	D	31	MECHANICAL PENCIL 0.5 MM BARREL	MECHANICAL PENCIL 0.5 MM BARREL
19	D	33	MECHANICAL PENCIL 0.5 MM BARREL	MECHANICAL PENCIL 0.5 MM BARREL
20	D	43	MECHANICAL PENCIL 0.5 MM TRANSPARENT VIOLET	MECHANICAL PENCIL 0.5 MM TRANSPARENT VIOLET

Source Text: 0.5 MM BURGUNDY MECHANICAL PENCIL  
Translated Text: 0.5 MM BURGUNDY MECHANICAL PENCIL

Quality Metrics: Quality: 0 - Parsed %: 100, In Dictionary %: 0 (Acceptable: 0, avg quality=0, Unacceptable: 34, avg quality=0.)

### サンプル・データ表

この表の複数の列は、「Test Translated Attributes」サブタブ（6-9 ページの「「Test Translated Attributes」サブタブ」を参照）の列と同じで、同じように動作します。異なる列は次のとおりです。

#### 長さ (Len)

元のテキストの文字列の長さを示します。

#### English, US

標準化の元のテキストのソース言語バージョンを示します。

#### 翻訳言語、国

元のテキストの翻訳済バージョンを示します。翻訳言語の名前およびその発祥国が列ラベルになります。

### ソース・フィールド

このフィールドには、ソース言語テキストが表示されます。このフィールドは編集可能で、[Enter] キーを押すと、データ・レンズ上で即時に結果をレビューできます。

### 「Translated」フィールド

このフィールドには翻訳済テキストが含まれ、「Source」フィールドに加えられた変更が表示されます。

### Quality Metrics

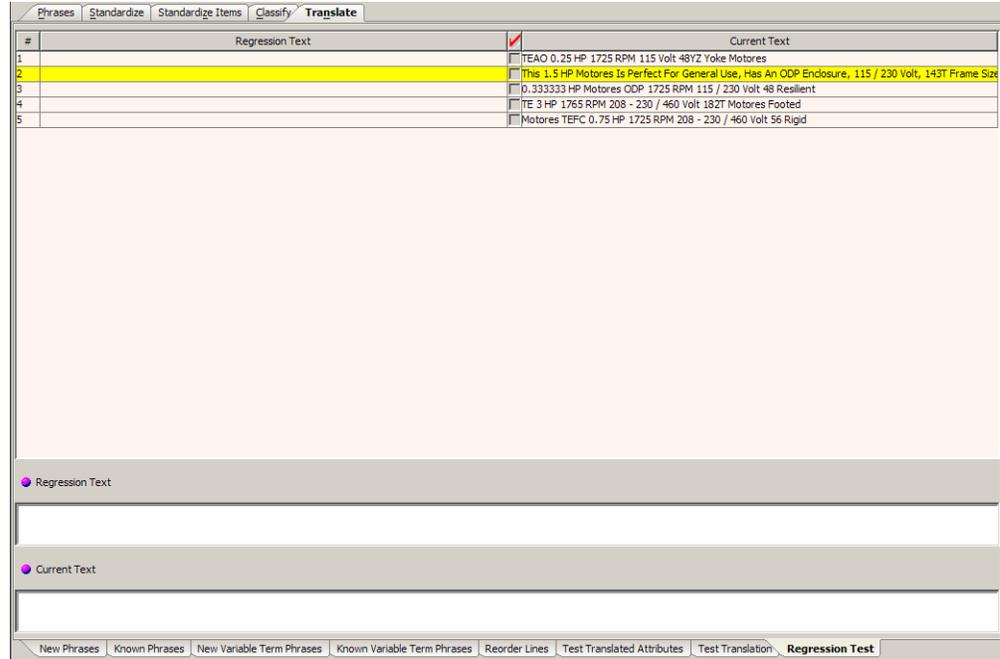
翻訳の進行状況を監視できるように、正確な翻訳に関する基準がレビュー用に表示され、次の内容が表示されます。

- 解析の品質
- デクシオナリで検出されたテキストの割合
- 全体的な品質インジケータの割合: 許容、平均および非許容

## 「Regression Test」サブタブ

回帰テストはテキスト翻訳の重要な部分で、翻訳が予想どおりであることを確認できます。

タブがアクティブでない場合は、「**Regression Testing Active**」データ・レンズ・オプションを設定します。詳細は、2-28 ページの「[回帰テストがアクティブであることの確認](#)」を参照してください。



回帰テストには、サンプル・データの「前」の状態と「後」の状態を示す2つのペインがあります。

### 「Before」ペイン

このペインには、回帰テスト前にデータ・レンズに定義されていたルールに基づいて翻訳されたデータが表示されます。このペインの選択したデータ行に表示されるテキストは、「**Current Text**」フィールドにも表示されます。

## 「After」 ペイン

このペインには、データ・レンズに定義されたルールに基づいて翻訳されたテキストが表示されます。このペインの選択したデータ行に表示されるテキストは、「**Regression Text**」フィールドにも表示されます。

「**Before**」 ペインおよび「**After**」 ペインにデータが表示されない場合は、サンプル・データが初期化されていないため、回帰ベースは存在しません。回帰ベースの初期化の詳細は、6-14 ページの「[回帰ベースの作成と更新](#)」を参照してください。

「**Before**」 ペインまたは「**After**」 ペインの色付けされたテキストは、次の内容を示します。

### 赤

削除されたテキスト。

### 緑

追加されたデータ。すべてのテキストは問題についてレビューする必要があるため、左右のペインを目視で比較します。

### オレンジ

この用語には該当する翻訳が適用されているため、回帰テキストと現在のテキストの両方が色付けされます。

## 「Review」 列

「**Review**」 列（赤色のチェック・マーク）は、新規または変更されたデータ行を示し、この行のテキストをレビューする必要があります。「**Current Text**」 列の情報が正しく、有効な推移として変更を受け入れる場合は、このチェック・ボックスを選択して、データが回帰テストに含まれるようにします。

## 回帰ベースの作成と更新

回帰ベースの作成におけるベスト・プラクティスは、サンプル・データを1つのファイルに結合することです。詳細は、6-14 ページの「[回帰ベースの作成と更新](#)」を参照してください。

ファイルの結合によってデータが削除されることはなく、選択したサンプル・ファイルが新規の大規模なファイルに単純に結合されます。

次に、回帰ベースのサンプル・データ・ファイルに対して1つの変更を加え、回帰セットを確認して適切に更新します。複数の変更を加えると回帰の結果が読みにくくなり、エラーを見落としたり修正が困難になる可能性があります。

回帰ベースを作成するには、「**File**」メニューの「**Create New Regression Base**」オプションを選択し、回帰テストに使用するサンプル・データ・ファイルを選択します。これによって、回帰ベースが初期化され、「**After**」ペインに結果が表示されます。

回帰ベースは、「**File**」メニューの「**Update Regression Base**」オプションを使用して、レビュー済および受入済のテキスト行（前述の「**Review**」列）で更新できます。

---

---

**注意：** サンプル・データをレビューまたは受け入れた場合のみ、回帰ベースを初期化または更新してください。

---

---

## 翻訳修正書式

翻訳修正では、次の機械翻訳の調整に全行翻訳を使用できます。翻訳修正は通常、機械翻訳による全行翻訳に挿入されている不要なテキストを削除するために使用します。

たとえば、「la Red」ではなく「Red」にして、一定の項目を削除します。データの複数の場所に目的に沿わない「la」が表示される可能性があります。このような問題がデータ内で繰り返し検出される場合は、**翻訳修正書式設定**機能を使用する必要があります。かわりに、翻訳ディクショナリにフレーズ翻訳を指定することもできます。

次の内容を考慮してください。

- 修正は行レベルでのみ機能します。
- 翻訳修正書式は、最初にコンサルタントの指導を受けてから適用してください。
- 「**Translation Repair Formats**」メニュー項目を使用できるのは、データ・レンズ翻訳機能を購入して、Oracle DataLens Server にインストールしている場合です。

この機能を使用するには、「**Data Lens**」メニューから「**Translation Repair Formats**」を選択します。

これによって、「**Edit Translation Repair Formats**」ダイアログがオープンします。このダイアログには、全行翻訳の後に実行するすべての置換内容がリストされています。このダイアログでは、各置換が1行ごとに表示されます。これらの置換は書式ルールと呼ばれます。

たとえば、書式ルールの「s/¥sEL¥s/ /gi」は4つの部分で構成されています。

- s/ で示された検索操作を実行します。
- 空白、EL、別の空白の順に並ぶテキスト文字列を識別します。空白の例には、タブやスペースがあります。
- 識別した文字列を空白で置換します。
- 最後のフォワード・スラッシュの後にある「gi」は置換修飾子です。「g」は、行内で置換をグローバルに実行することを指定し、この修飾子がない場合は最初のインスタンスのみが置換されます。「i」は大 / 小文字の区別を無視します。したがって、「El」や「el」も含めて「EL」のすべての書式は空白に置換されます。

---

**注意：** パターン検索および置換の詳細は、[付録 B 「正規表現」](#) を参照してください。

---

置換の結果のテストは重要なため、「**Test Translation**」サブタブに戻り、サンプル・コンテンツを再度翻訳してください。データをフィルタ処理し、目的の翻訳が実行されていることを確認します。



# 7

---

---

## レポートの生成

この章では、Knowledge Studio のレポート作成機能について説明します。

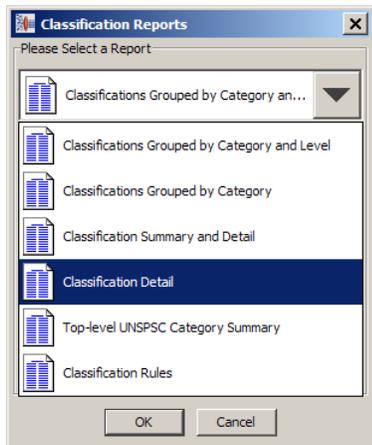
## 機能固有のレポート

これらのレポートには、Knowledge Studio 内の特定の機能に関する情報が表示されます。

### Classification Reports

分類レポートは、結果のテストおよび重要な分類統計のレコードの表示に役立ちます。

「Classify」タブから「Test Classification」サブタブを選択します。このレポートがアクティブになるのは、この場所からのみです。「File」メニューから「Report...」をクリックします。



生成するレポートを選択し、「OK」をクリックします。

このレポートは、次のサンプルで示すようにデフォルト・ブラウザに表示されます。

#### Classifications by Category and Level (UNSPSC/eClass)

DataLens: Writing\_Instruments\_Tutorial\_abc  
 Classification Type: UNSPSC\_WRITING\_INST  
 Sample Data File: 1-sample01.xml  
 Reported: Fri Dec 12 16:26:47 2008

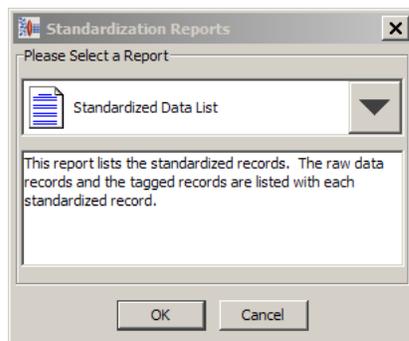
Parent ID	Category ID	Category Name	Level	Product ID	Product Name
	00000000	Unclassified			
			Level 1 (Segment)		Test Item Description Typed Into The Selected Line Page
44121700	44121702	Pen or pencil sets			
			Level 4 (Commodity) W16		Ballpoint Pen Medium Point Black Barrel Black Ink
			Level 4 (Commodity) W117		Comfort Grip Retractable Ballpoint Pen Medium Point Blue Ink
			Level 4 (Commodity) W118		Comfort Grip Retractable Ballpoint Pen Medium Point Red Ink
			Level 4 (Commodity) W12*		Retractable Ballpoint Pen Non - Refillable Black Ink
			Level 4 (Commodity) W128		Erasable Ink Ballpoint Pen Medium Point Blue Barrel Char Slash Blue Ink

分類レポートは、ブラウザから保存または印刷できます。

## Standardization Report

標準化レポートを使用すると、現在のサンプル・データ・ファイルの標準化の結果に関する標準化レポートを表示できます。

「**Standardization**」タブから「**Test Global Standardization**」サブタブを選択します。このレポートがアクティブになるのは、この場所からのみです。「**File**」メニューから「**Report...**」をクリックします。



「**Standardized Data List**」を選択し、「**OK**」をクリックします。

このレポートは、次のサンプルで示すようにデフォルト・ブラウザに表示されます。

### Standardized Data List

DataLens: Writing\_Instruments\_Tutorial\_abc  
 Sample Data File: 1-sample01.xml  
 Reported: Fri Dec 12 16:15:40 2008

---

ID, Ambiguity Count, Quality / Raw Text / Standardized Text (Bold) / Tagged Text

---

ID = W169, Ambiguity Count = 1, Quality = 100  
 .5MM Burgundy Barrel Mechanical Pencil  
**.5 Millimeter Burgundy Barrel Mechanical Pencil**  
 [a\_size] (.5 Millimeter) [a\_barrel\_color] (Burgundy Barrel) [a\_pencil\_type] (Mechanical) [i\_pencil] (Pencil)

---

ID = W141, Ambiguity Count = 1, Quality = 86  
 Automatic Pencil .5MM Kryptonite Green Barrel  
**Automatic Pencil .5 Millimeter Kryptonite Green Barrel**  
 [a\_pencil\_type] ( Automatic) [i\_pencil] ( Pencil) [a\_size] (.5 Millimeter) Kryptonite [a\_barrel\_color] ( Green Barrel)

---

ID = W139, Ambiguity Count = 1, Quality = 55  
 Automatic Pencil with Eraser Metal Clip .5MM Black Cap  
**Automatic Pencil With Eraser Char Slash Metal Clip .5 Millimeter Black Cap**  
 [a\_pencil\_type] ( Automatic) [i\_pencil] ( Pencil) With Eraser [i\_ignored\_punctuation] ( Char Slash) Metal Clip [a\_size] (.5 Millimeter) [a\_hue] ( Black) Cap

---

標準化レポートは、ブラウザから保存または印刷できます。

前述のレポートの各行の品質指標 (QI) は、データ・レンズで認識された行の割合です。すべての用語ルールを「**Standardize**」タブでレビューおよび標準化している場合、この数値は標準化テキストの量と密接に関連している場合があります。

## Grammar Reports

これらのレポートには、Knowledge Studio 内で定義されたルールの文法に関する情報が表示されます。

「Phrases」タブから「Define Phrases」サブタブを選択します。このレポートがアクティブになるのは、この場所からのみです。「File」メニューから「Report...」をクリックします。



次のレポートから必要なレポートを選択します。

### Grammar Rules

このレポートではすべてのルールが表示され、最初にフレーズ、次に用語が、関連する生成結果とともに表示されます。

### Grammar Rules by Domain

このレポートではすべてのルールがドメイン別にグループ化されて表示され、最初にフレーズ、次に用語が、生成結果とともに表示されます。

このレポートは、次のサンプルで示すようにデフォルト・ブラウザに表示されます。

## Grammar Rules

DataLens: MRO\_Demo\_Passives  
Reported: Thu Dec 11 09:41:11 2008  
User: dleeper

---

### Phrases

```
[a_3_dimensions]
[a_dimension_length][char_x][a_dimension_length][char_x][a_dimension_length]
[number][char_x][a_dimension_length][char_x][a_dimension_length]
[number][char_x][number][char_x][a_dimension_length]
[number][char_x][number][char_x][number]

[a_additional_data]
[additional_data]

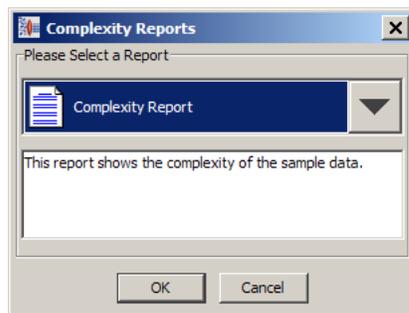
[a_adjustable]
[adjustable]

[a_aluminum_snap_in]
[a_material][snap_in]
```

文法レポートは、ブラウザから保存または印刷できます。

## Complexity Reports

このレポートでは、サンプル・データの複雑度が表示されます。「File」メニューから「Complexity Reports...」をクリックします。



このレポートは、次のサンプルで示すようにデフォルト・ブラウザに表示されます。

### Data Metrics for MRO\_Demo\_Passives

DataLens - MRO\_Demo\_Passives  
 DataLens Path - C:\Documents and Settings\dleeper\Application Data\DataLens\data\project\MRO\_Demo\_Passives  
 Sample Data File - Res\_chip\_arrays-sample01.xml (368 lines, 92.30% parsed)  
 Level-of-Effort Data File Collection - Res\_chip\_arrays-sample\*.xml (1 files)  
 Text Standardization - Match\_Attributes  
 Unit Conversion - Default  
 Report Date/Time - Thu Dec 11 09:44:29 2008

Metrics appear in bold and headline each table. Values range from 0 (worst) to 100 (best). E.g., The least "intelligible" or most "cryptic" will have a score of 0, and the least canonical will have a score of 0, while the closest to standard will have a score of 100. Additional supporting information appears below each summary metric score.

#### Standardization Level

Measures the difference between standardized text and raw input text for the selected file. The score will be artificially high if the DataLens is incomplete (unknown text or incomplete standardization rules). Term standardization coverage is 53.23%. Phrase standardization coverage is 2.97%.

Estimated Standardization Level (out of 100)	<b>80.72</b>
Difference Between Input and Standardized Text	19.64%

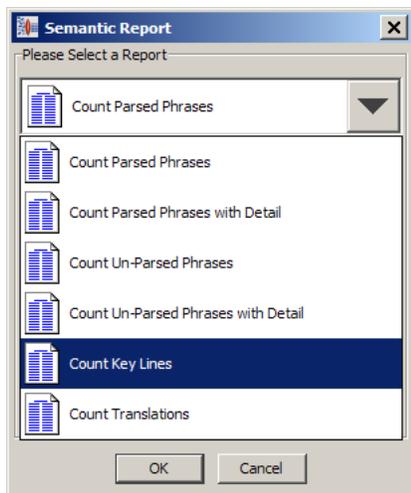
#### Data Useability

Data useability combines intelligibility and consistency. Consistency for query simplicity, key integrity, and join completeness. Intelligibility for query validity, non-expert understanding, and external sharing. Here useability dimensions for this DataLens are compared with a variety of other DataLenses.

複雑度レポートは、ブラウザから保存または印刷できます。

## Semantic Reports

これらのレポートでは、任意のツールで利用できるテキスト・ファイルを作成し、表示された情報をレビューします。セマンティック・レポートは、レポートの生成時に選択できるサンプル・ファイルに基づいています。「File」メニューから「Semantic Reports...」をクリックします。



次のレポートから必要なレポートを選択します。

### Count Parsed Phrases

このテキスト・ファイルでは、データ・レンズで解析したフレーズを、選択したサンプル・ファイルに使用した回数をカウントします。これには、フレーズの使用頻度、関連するフレーズおよび関連する用語が表示されます。

### Count Parsed Phrases with Detail

このレポートでは、データ・レンズで解析したフレーズを、選択したサンプル・ファイルに使用した回数をカウントします。これには、フレーズの使用頻度、関連するフレーズおよび関連する用語の詳細が表示されます。

### Count Un-Parsed Phrases

このレポートでは、データ・レンズで未解析のフレーズを、選択したサンプル・ファイルに使用した回数をカウントします。

### Count Un-Parsed Phrases with Detail

このレポートでは、データ・レンズで未解析のフレーズを、選択したサンプル・ファイルに使用した回数をカウントします。

### Count Key Lines

このレポートでは、プロジェクトのキー行の数をカウントします。

### Count Translations

このレポートでは、プロジェクトの翻訳済の行の数をカウントします。

目的のサンプル・ファイルを選択します。このレポートは、生成された後、「Status」ペインに表示されるパスのテキスト・ファイルに保存されます。

---

## データ・レンズおよびファイルの管理

この章では、データ・レンズおよびサンプル・データ・ファイルの管理に関連する様々なタスクについて説明します。

## データ・レンズの管理

この項では、データ・レンズの管理方法について説明します。

### データ・レンズ・オプションの設定

データ・レンズの操作に影響を与える様々なデータ・レンズ・オプションは、ユーザーが構成できます。「Data Lens」メニューから「Data Lens Options」を選択します。

「General」タブは、データ・レンズ・オプションを設定する際のデフォルトです。

#### Classification Context

Knowledge Studio に対して、分類スキーマ内でのドラッグ・アンド・ドロップ機能では分類の親に子を含めるように指示します。

#### Case Sensitive Text Parsing

このオプションを設定した後は、大 / 小文字の区別が保持され、用語ノード（紫色）のみに適用されます。用語ノードには、特定のテキスト品目のすべてのバリエーションが含まれます。たとえば、「Cap」および「CAP」は個別に指定する必要があります。このチェック・ボックスの選択を解除すると、大文字表記での差異は無視されるため、「cap」の指定は 1 回のみ必要です。このオプションを切り替えると、すべての用語およびフレーズのルールが新しい大 / 小文字区別モードで統一されます。

#### Full Form/Generate Term Variants

用語ルールの作成時に、用語バリエーションが自動的に生成されます。たとえば、「RES」は「RESISTOR」の略語です。

#### Generate Phrase Variants

標準化された出力から、用語およびフレーズの生成結果が自動的に作成されます。

#### Apply Button also Refreshes sample data display

「Apply」ツールバー・ボタンを、インタフェース・ペインをリフレッシュし、編集による変更を現在オープンしているサンプル・ファイルに適用するように変更します。

#### Regression Testing Active

回帰テストのタブが Knowledge Studio 全体でアクティブになります。

#### View Hierarchy

すべての階層構造が完全に表示されます。

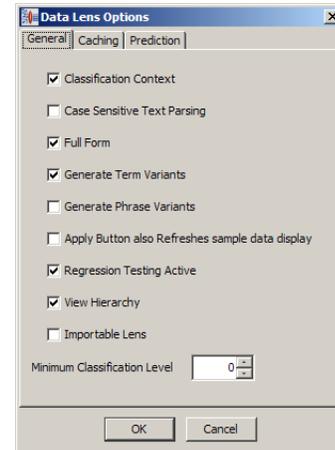
#### Importable Lens

Oracle DataLens Server にチェックインした後のデータ・レンズを他のデータ・レンズにインポートできます。このようなデータ・レンズは**基盤データ・レンズ**と呼ばれます。

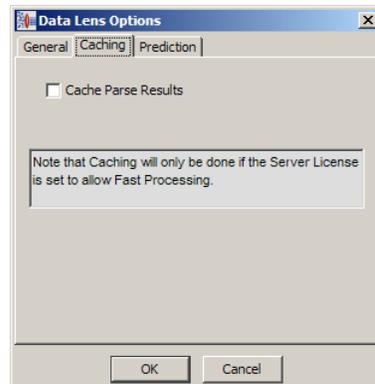
**注意:** この機能をアクティブ化するには、コンサルタントに問い合わせる必要があります。

#### Minimum Classification Level

データ・レンズ内で定義できる分類の最小レベルを決定できます。最大レベルは 5 です。



「Caching」タブを選択すると、データ・レンズの解析データ結果のキャッシュを設定できます。



選択すると、各データ行の解析後にその結果がグローバル・キャッシュにキャッシュされるため、データ・レンズによる再解析が不要となり、データの処理速度が向上します。

「Prediction」タブを選択すると、用語予測機能に関連するオプションを設定できます。

#### Prediction Threshold (%)

フレーズ・ルールの信頼度しきい値を0（ゼロ）から100%の範囲で設定できます。このしきい値は、認識されていない入力テキストが既存の用語と一致しているかどうかを判断し、予測するために使用されます。

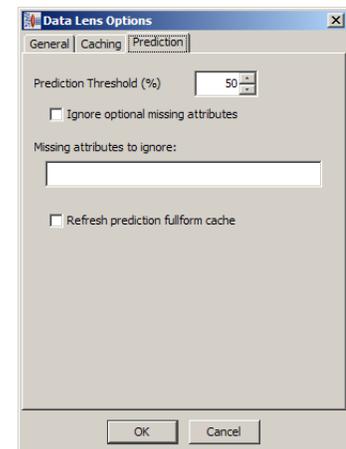
**注意：**このオプションを0（ゼロ）に設定すると誤った結果が生成される場合があるため、設定する前に予測を慎重に検討する必要があります。

#### Ignore optional missing attributes

欠落している属性を予測機能が無視するように指定できます。

#### Missing attributes to ignore:

予測機能が無視する属性を指定できます。



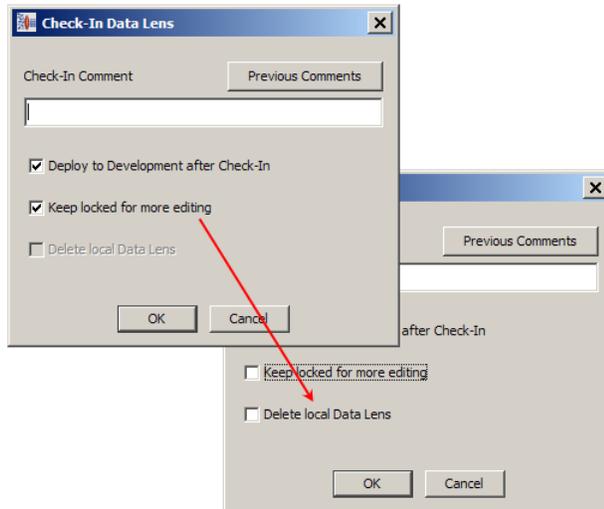
#### Refresh prediction fullform cache

「OK」を選択すると、予測機能で使用するために保持されている完全形式の用語バリエーションのキャッシュがリフレッシュされます。このオプションは使用後にリセットされます。

## データ・レンズのチェックイン

チェックインできるのは、チェックインしていないデータ・レンズか、現在のユーザーが以前チェックアウトしたデータ・レンズのみです。チェックインしたデータ・レンズは、読取り専用に変更されます。

1. 「Data Lens」メニューから「Check-in Data Lens...」をクリックします。



2. 新しいリビジョンに関連付けるチェックイン・コメントを入力します。「**Previous Comments**」ボタンを使用すると、新しいチェックイン・コメントで使用するために、他のコメントのテキストを表示してコピーできます。

次の機能を持つチェック・ボックスから選択します。

- **Deploy to Development after Check-In**

チェックインしたデータ・レンズは他のユーザーが使用できますが、DSA で使用する場合は使用しているバージョンが更新されます。

- **Keep locked for more editing**

このデータ・レンズは、他のユーザーが編集のためにチェックアウトすることはできません。このチェック・ボックスの選択を解除すると、データ・レンズは読取り専用となり、「Delete local Data Lens」チェック・ボックスがアクティブになります。

- **Delete local Data Lens**

このデータ・レンズをローカル・ドライブから削除します。

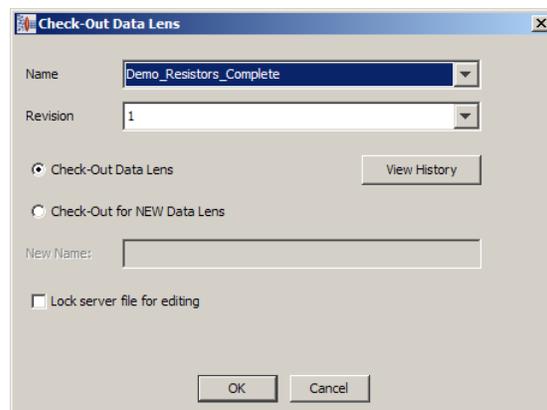
3. 「OK」をクリックします。

データ・レンズをクローズしたことの確認が表示され、データ・レンズ自体はクローズした状態になります。

## データ・レンズのチェックアウト

チェックアウトできるのは、現在のユーザーがチェックインしたデータ・レンズのみです。

1. 「Data Lens」メニューから「Check-Out Data Lens...」をクリックします。



2. 使用可能なデータ・レンズのリストからチェックアウトするデータ・レンズを選択します。  
「Revision」は変更不可で、自動的に増分されます。
3. 選択したデータ・レンズのチェックイン / チェックアウト履歴をレビューする場合は、「View History」をクリックします。「Revision History」ダイアログが表示され、データ・レンズの各リビジョンに関するすべてのコメントがリスト表示されます。
4. 2つのオプションを使用して、選択したデータ・レンズを単にチェックアウトするか、新しいデータ・レンズを選択のベースにするかを選択します。「Check-Out for NEW Data Lens」を選択した場合は、表示されているフィールドに新しい名前を指定する必要があります。
5. データ・レンズはチェックアウトするが、編集はしない場合は、「Lock server file for editing」を選択してデータ・レンズを読取り専用にします。
6. 「OK」をクリックします。  
データ・レンズをクローズしたことの確認が表示され、「読取り専用」が表示されます。

## データ・レンズの削除

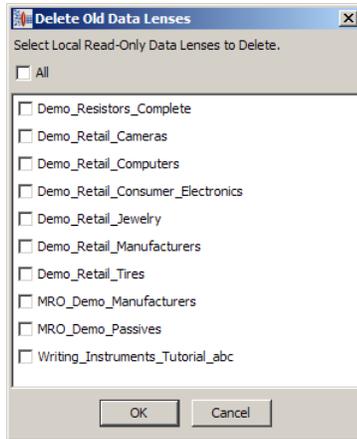
ローカル・コンピュータ上の既存のデータ・レンズは、Knowledge Studio 内から削除できません。

1. 削除するデータ・レンズをオープンします。
2. 「Data Lens」メニューから「Delete Data Lens」をクリックします。
3. この処理に対する警告が表示されます。
4. データ・レンズを削除するには、「OK」を、保持するには、「Cancel」をクリックします。

## 読取り専用のデータ・レンズの削除

また、ローカル・コンピュータにある読取り専用（チェックイン）バージョンのデータ・レンズも削除できます。

1. 「File」メニューから「Delete Read-Only Lenses」をクリックします。

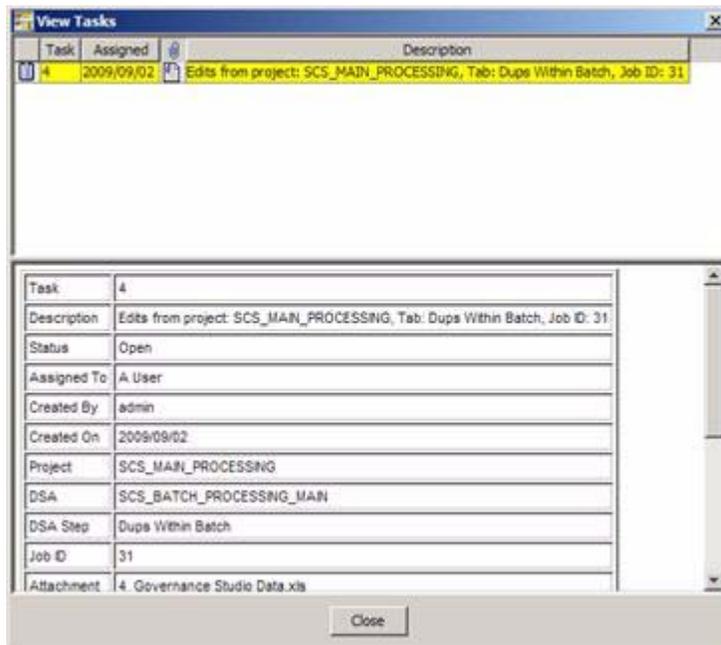


2. 「All」を選択するか、削除する個別のデータ・レンズを選択します。
3. 「OK」をクリックし、選択したデータ・レンズを削除します。  
選択したデータ・レンズがローカル・ドライブから削除されます。

## タスクの表示

この機能で使用するように割り当てられたタスクがあるかどうかを確認できます。

- 「View」メニューから「View My Tasks」を選択します。



上部ペインにすべての割り当て済タスクが表示され、下部ペインには、選択したタスクの詳細が表示されます。

---

**注意：** 下部ペインに表示されるフィールドは編集可能ですが、変更内容は保存されません。

---

### Change Task Status

タスクのステータスの変更については、8-7 ページの「[タスク・ステータスの変更](#)」を参照してください。

	Change Task Status	s
	Download Attachments	d
	Create Task	c

### Download Attachments

タスクの作成時に保存されたファイルをダウンロードして、タスクの完了に使用できます。ファイルの保存先ディレクトリを選択するための、ファイルの保存ダイアログが表示されます。

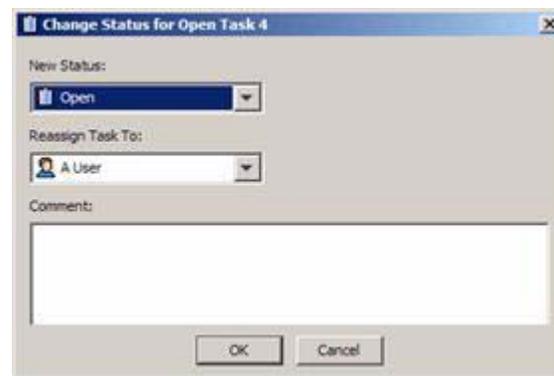
### Create Tasks

新規タスクの作成については、8-8 ページの「[タスクの作成](#)」を参照してください。

上部ペインのコンテキスト依存メニューは、添付アイコンを右クリックするとアクティブ化され、次のように使用されます。

## タスク・ステータスの変更

このオプションを選択すると、タスクのステータスを変更したり、タスクを他のユーザーに再割当てすることができます。



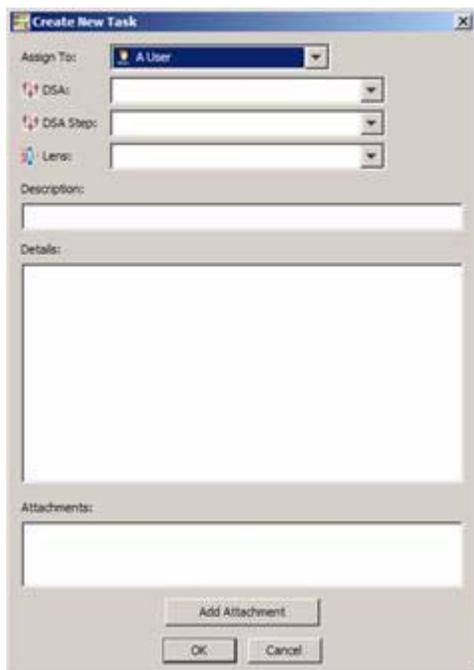
1. 新しいステータスまたはタスクを再割当てするユーザー（あるいはその両方）をリストから選択します。

**ヒント：** このタスクのレビュー担当者が不明なため、適切な担当者を後で割り当てる場合は、「Unassigned Tasks」ユーザーを使用できます。

2. 将来の参照のために変更に影響を与えた理由を反映したコメント、または担当となった理由をタスクの新しい担当者に通知するためのコメントを入力します。
3. 「OK」をクリックします。

## タスクの作成

このオプションを選択すると、完全に新しいレビュー・タスクを作成できます。



1. このタスクを完了するユーザーを選択します。
2. 変更する DSA および DSA ステップを選択します。
3. 変更を適用するデータ・レンズを選択します。
4. タスクの実行方法に関する説明および特別な指示を入力します。
5. 添付するデータ・ファイルまたは他の情報がある場合は、「Add Attachment」をクリックしてファイルの場所を特定し、「OK」をクリックします。

必要なすべてのファイルを添付するまで、このステップを繰り返します。

6. 「OK」をクリックします。

タスクが作成され、そのタスクの詳細が記載された電子メールが割り当てたユーザーに送信されます。

## データ・レンズに関する情報

この項では、データ・レンズに関連する情報の表示方法と変更方法について説明します。

### データ・レンズの摘要の編集

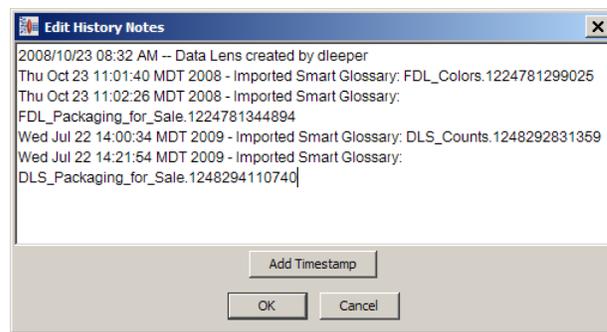
データ・レンズには、そのデータ・レンズの詳細な摘要を指定できます。

「Edit」メニューから「Edit Lens Description」をクリックします。データ・レンズに割り当てる摘要を入力し、「OK」をクリックします。

### データ・レンズの履歴ノート of 編集

この機能では、データ・レンズに関する情報履歴が提供されるため、今までに実行された保守を追跡できます。

1. 「Edit」メニューから「Edit Lens Description...」をクリックします。



2. 既存のリストに独自の情報を入力できます。
3. リストの各エントリにタイムスタンプを割り当てるには、エントリの末尾にカーソルを置き、「Add Timestamp」をクリックします。
4. タイムスタンプも含めてすべてのエントリは変更可能です。
5. 「OK」をクリックし、履歴ノートを保存します。

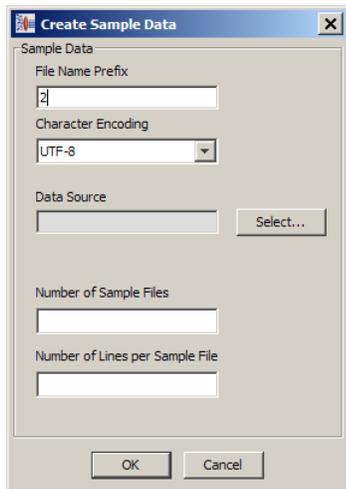
## サンプル・ファイル

この項では、データ・レンズのサンプル・データ・ファイルの使用方法について説明します。

### 新しいサンプル・データの作成

この機能を使用すると、新しいサンプル・データ・ファイルを作成して既存のセットに追加できます。

1. 「Edit」メニューから「New Sample Data」を選択します。



2. デフォルトの数字のファイル名接頭辞を使用するか、命名規則を入力します。

この名前は、新しいデータ・レンズの作成時に作成したファイルに類似していますが、特定の接頭辞が含まれます。たとえば、デフォルトの数字接頭辞を使用すると、初めてサンプル・データをインポートしたときのファイルの名前は「2-baseline.xml」や「2-sample1.xml」などになります。同様に、次回サンプル・データをインポートしたときのファイルの名前は「3-baseline.xml」や「3-sample1.xml」などになります。

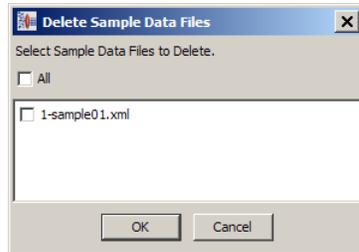
3. 適切なタイプの文字エンコーディングをリストから選択します。
4. 「Select...」をクリックしてポップアップ・メニューから「Data Source」タイプを選択し、使用するファイルを検索します。
5. 各ファイルに含めるサンプル・ファイルおよび行の数を入力します。
6. 「OK」をクリックします。

新しいサンプル・ファイルが、使用しているデータ・レンズのデータ・サブディレクトリに作成されます。

## サンプル・ファイルの削除

このオプションを使用すると、作成してデータ・レンズに関連付けたサンプル・ファイルを削除できます。

1. 「File」メニューから「Delete Sample Files...」をクリックします。

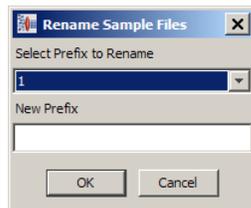


2. 削除するすべてのサンプル・ファイルまたは特定のサンプル・ファイルを選択します。
3. 「OK」をクリックし、選択したサンプル・ファイルを削除します。

## サンプル・ファイルの名前の変更

これを使用するのは、識別を容易にするためにサンプル・ファイルのグループの名前を変更する場合です。

1. 「File」メニューから「Rename Sample Files...」をクリックします。



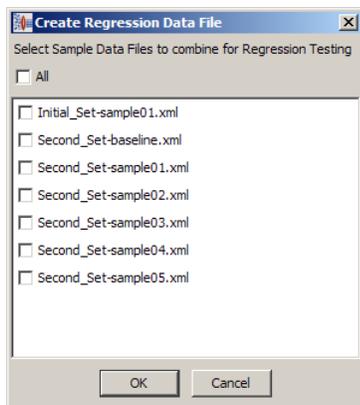
2. 名前を変更するサンプル・ファイル接頭辞をリストから選択します。
3. 使用する新しいファイル接頭辞を入力します。
4. 「OK」をクリックします。

サンプル・ファイルの名前の変更をレビューするには、「File」メニューから「Select Data File...」をクリックし、表示されたリストをレビューします。

## サンプル・ファイルの結合

この機能は、既存のサンプル・ファイルを組み合わせる新しい回帰サンプル・ファイルを作成する必要がある場合に使用できます。元のサンプル・ファイルは変更または削除されません。

1. 「File」メニューから「Combine sample data」をクリックします。



2. 結合するサンプル・ファイルを選択し、「OK」をクリックします。
3. 新しいサンプル・ファイルの名前を入力します。

新たに作成したファイルを開く場合は、「File」メニューから「Select Data File...」をクリックします。新しいファイルを検索して開きます。

---

## エクスポートとインポート

この章では、データ・レンズ、フレーズおよび用語ルール、および品目定義のインポートおよびエクスポート方法について説明します。

## データ・レンズのエクスポート

データ・レンズはエクスポートが可能で、アーカイブおよびバックアップ目的に対して有用です。

データ・レンズをエクスポートするには、「File」メニューの「Export Data Lens」をクリックします。データ・レンズのわかりやすいファイル名を入力し、「OK」をクリックします。

エクスポートしたすべてのデータ・レンズは、エクスポート・ディレクトリの `...¥Application%Data¥export` に格納されます。

## データ・レンズのインポート

エクスポートしたデータ・レンズを使用するにはインポートします。

データ・レンズをインポートするには、「File」メニューの「Import Data Lens」をクリックします。データ・レンズを検索してダブルクリックし、インポート・プロセスを起動します。

Knowledge Studio では、このデータ・レンズがローカルに存在するかどうかを確認し、存在する場合は、ファイルの上書きを回避するためにデータ・レンズの名前を変更するようにプロンプトが表示されます。

---

---

**注意：** データ・レンズをアプリケーションで使用するには、Oracle DataLens Server にチェックインする必要があります。

---

---

## ルールのエクスポート

レンズ・ルールはテキスト・ファイルにエクスポートでき、この出力はルールをレビューしたり、既存のルール・セットを前のルール・セットと比較する際に役立ちます。

データ・レンズのあるバージョンをチェックアウトしてテキスト・ファイルを作成し、次に同じデータ・レンズのそれより前のバージョンをチェックアウトします。次に比較を生成できる任意のツールを使用して、ファイル間の相違を比較できます。この機能は、データ・レンズを保守していて結果が予想どおりでない場合に使用します。

ルールは、データ・レンズ全体またはドメイン別にエクスポートできます。

ルールはデフォルトのエクスポート・ディレクトリの `...¥data lens name¥Application Data¥DataLens¥export` にエクスポートされますが、パスとファイル名を指定することもできます。

---

---

**注意：** データ・レンズ・ルールのエクスポートはルールのインポートの反対ではなく、エクスポートしたルールはインポート可能な形式ではありません。エクスポートしたルールのインポートについては、コンサルタントにお問い合わせください。

---

---

## フレーズおよび用語のインポート

ユーザーの組織内に存在するナレッジは、データ・レンズを作成および補完する際の使用、フレーズおよび標準化のナレッジ構築の一部として使用できます。この情報は、「**Import Phrases and Terms**」機能を使用して、Excel スプレッドシートからデータ・レンズにインポートできます。

Excel スプレッドシートまたはタブ区切りファイル（CSV またはテキスト）の形式を選択できます。たとえば、フレーズ、フレーズの使用および初期データを示す列を含めることができます。

次の各例では、列の作成方法およびスプレッドシート・ファイルの移入方法について説明します。

	A	B	C
1	phrase	term	text
2			

	A	B	C	D
1	phrase	term	regex	Default
2				

用語バリエントの作成に使用できる列は、次のように 2 つあります。

### fullform

完全形式の用語ルールの変種をインポートできます。

### handling

用語バリエントを作成するためのオプションを使用できます。

---

**注意：**「fullform」および「handling」列は、スプレッドシートの「text」または「regex」列の後で、任意の標準化列の前に配置する必要があります。

---

「fullform」および「handling」列で使用できる値は、次のとおりです。

### Expand

テキスト・ノードの用語バリエントが自動的に作成されます。

### Plural

用語バリエントおよび複数形が自動的に作成されます。

### なし（空白セル）

用語バリエントは作成されず、用語の変更はありません。

	A	B	C	D	E
1	phrase	term	text	fullform	handling
2	a_marker_type	highlighter	hi-lighter	highlighter	expand
3	a_marker_type	highlighter	hi-liters	highlighter	expand
4	a_marker_type	highlighter	Highliger	highlighter	expand
5	a_marker_type	highlighter	Hilighter	highlighter	expand
6	a_pen_type	roller_ball	roller	rollerball	expand
7	a_pen_type	roller_ball	roller ball	rollerball	expand
8	a_pen_type	roller_ball	rollerball	rollerball	expand
9	a_pen_type	roller_ball	rolleball	rollerball	expand

さらに、標準化ルールは同じスプレッドシート内にインポートできます。標準化タイプは「**Full Description**」および「**Short Description**」です。

---

**注意：** 特定の標準化内にルールをインポートするには、標準化タイプがデータ・レンズ内に存在する必要があります。

---

新しいナレッジをデータ・レンズにインポートする手順は、次のとおりです。

1. 「**File**」メニューから「**Import Phrases and Terms**」をクリックします。
2. インポート・スプレッドシートを参照し、「**Open**」をクリックします。

ルールがデータ・レンズにインポートされます。新たにインポートしたこれらのルールの追加によって、認識の割合が増加します。

## Excel を使用した、ルールおよびフレーズのインポート時の別名の作成

Excel を使用してルールおよびフレーズをインポートするときは、別名を作成できます。Excel スプレッドシートには「**Alias**」列がないため、別名は、属性または品目定義の名前に基づいて自動的に作成されます。Knowledge Studio では、ユーザーがスプレッドシートに指定したラベルを別名と想定し、必要な規則に従った属性名を作成するために、必要に応じてその名前を変更します（たとえば、属性名の空白をアンダースコアに変更するなど）。

これらの別名は、編集または表示できます。詳細は、2-17 ページの「**別名**」を参照してください。

## 品目定義のエクспортとインポート

データ・レンズ間での品目定義のインポートとエクспортによって、品目定義情報を新しい方法で再利用できます。

品目定義は、データ・レンズからエクспортして、別のデータ・レンズに簡単にインポートできます。この機能を使用すると、単一の品目定義、または品目定義階層の分岐全体を新規または既存のデータ・レンズにインポートできます。

品目定義をインポートすると、その品目定義に適用されているすべてのルールもインポートされます。

1. ソース・データ・レンズをオープンします。

これは、ソース・データ・レンズからエクспортしてターゲット・データ・レンズにインポートする品目定義（または品目定義の階層）が含まれているデータ・レンズです。

2. ソース・データ・レンズに含まれているナレッジを最大限に利用するために、インポートする品目定義およびすべての単位換算のナレッジに関連するすべてのスマート用語集がターゲット・データ・レンズに含まれていることを確認してください。

スマート用語集をインポートする必要がある場合は、13-2 ページの「**スマート用語集のインポート**」を参照してください。

3. 「**Phrases**」タブの「**Define Items**」サブタブを選択します。
4. エクспортする品目定義の分岐を右クリックし、「**Export Item Definition**」をクリックします。
5. エクспорт・ファイルの名前を入力するか、デフォルトのファイル名をそのまま使用して「**Save**」をクリックし、エクспортされた品目定義を保存します。
6. 保存先のデータ・レンズをオープンします。
7. 「**Phrases**」タブの「**Define Items**」サブタブを選択します。
8. インポートする品目定義の親分岐を右クリックし、「**Import Item Definition**」をクリックします。

9. 保存したエクスポート・ファイルを選択して「Open」をクリックし、品目定義をインポートします。

品目定義およびその品目定義に関連付けられているすべてのルールがインポート先のデータ・レンズにインポートされます。

## 品目定義のエクスポートとインポートに関するトラブルシューティング

この項では、発生する可能性があるいくつかの問題とその解決方法について説明します。

### 既存の品目定義と同じ名前のインポート品目定義

ターゲット・データ・レンズにすでに存在する品目定義はインポートできないため、インポートを試みると、インポートできないことを示す情報メッセージが表示されます。

最適な解決方法は、品目定義の名前を変更することです。

### スマート用語集に含まれるナレッジに依存するソース品目定義

スマート用語集に依存する品目定義（単位の単位換算など）をエクスポートした場合は、その品目定義をインポートする前に、ターゲット・データ・レンズにスマート用語集をインポートしてください。インポートしない場合は、一部のナレッジをターゲット・レンズ内に再作成することが必要になる場合があります。

### 標準化情報に依存するソース品目定義

品目定義をターゲット・データ・レンズにインポートする場合は、ターゲット品目定義に必要なすべての標準化タイプを作成してください。標準化情報は、品目定義のインポート時に保存されません。

### 単位換算情報に依存するソース品目定義

単位換算が必要な品目定義をインポートする場合は、ターゲット・データ・レンズにすべての単位換算ナレッジを作成してください。単位換算情報は、品目定義のインポート時に保存されません。

### 値ロジック情報に依存するソース品目定義

ターゲット・データ・レンズに必要な単位換算情報が設定されている場合は、インポートの後に属性値ロジックが保存されます。多くの場合、単位換算情報はDLS\_Units\_of\_MeasureやDLS\_Units\_of\_Measure\_Retailなどのスマート用語集に含まれるため、適切なスマート用語集をターゲット・データ・レンズにインポートすると、この問題は回避されます。

インポートした品目定義の属性で使用する検索ロジックは、インポートで保存されます。



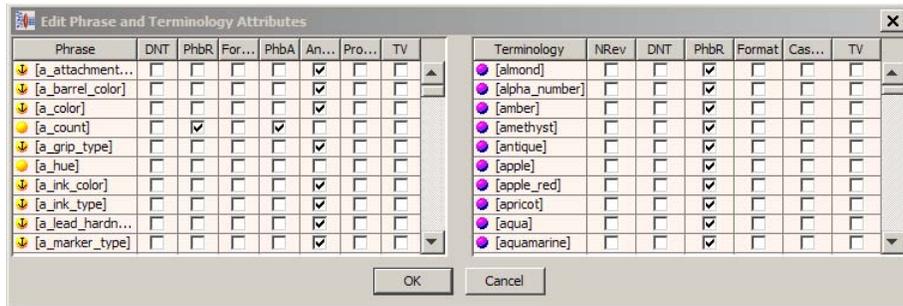
---

## コンテキストおよび品目定義の詳細な定義

この章では、Knowledge Studio を使用したコンテキストおよび品目定義の詳細な定義に関連する技術および情報について説明します。

## 属性の編集

「Edit」メニューから「Edit Phrase and Term Attributes」をクリックします。



次の各列のチェック・ボックスを使用して、フレーズおよび用語属性を編集します。

### Do Not Translate (DNT)

フレーズまたは用語ルールに対して「Do Not Translate」属性を設定 / 設定解除します。

### Prohibit Rename (PhbR)

選択したルールに対して「Prohibit Rename」機能を設定 / 設定解除します。選択したルールに対してこのチェック・ボックスを選択すると、そのルールの名前は変更できなくなります。

### Format to Locale

フレーズまたは用語ルールに対して「Format to Locale」属性を設定 / 設定解除します。

### Prohibit Anchor (PhbA)

選択したフレーズ・ルールに対して「Anchor」属性を設定 / 設定解除します。選択したルールに対してこのチェック・ボックスを選択すると、そのルールはフレーズ構造の上部に配置できなくなります。このルールを起動する場合は、親を指定する必要があります。

### Anchor

選択したフレーズ・ルールに対して「Anchor」属性を設定 / 設定解除します。このチェック・ボックスを選択すると、選択したフレーズ・ルールはフレーズ構造の上部に配置されて親ノードが不要になります。

### Promote Children

フレーズ・ルールに対して「Promote Children」属性を設定 / 設定解除します。これによって下位レベルのフレーズを個別に昇格できます。

### Translate Variable (TV)

フレーズおよび用語ルールに対して「Translate Variable」属性を設定 / 設定解除します。これによって、このフレーズまたは用語（あるいはその両方）を、それらが使用されるコンテキストの場所に関係なく 1 回で翻訳できます。

### Needs Review (NRRev)

用語ルールに対して「Needs Review」属性を設定 / 設定解除します。

### Case Sensitive

用語ルールに対して「Case Sensitive」属性を設定 / 設定解除します。

属性を複数行の範囲で設定または設定解除するには、対象となる最初のルールの属性をクリックし、[Shift] キーを押しながら最後のルールの属性をクリックします。これは、単一の属性列内でのみ機能します。

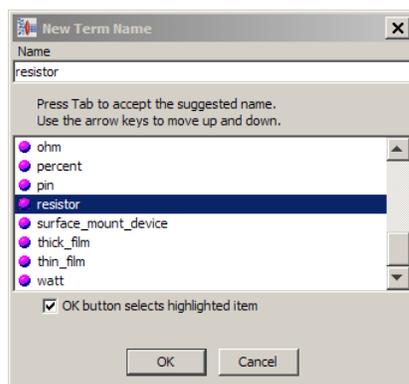
ルール名をダブルクリックすると、そのルールに対する「Review Productions」ダイアログがオープンします。

## 用語バリエーションの生成

用語ルールを定義すると、Knowledge Studio では、該当する用語の略語を自動的に生成できます。たとえば、[resistor] 用語ルールを設定すると、Knowledge Studio では、rsstr など様々な代替を自動的に生成できます。多くの場合は、これによって、追加の関連付けを作成するための手動によるドラッグ・アンド・ドロップが不要になります。

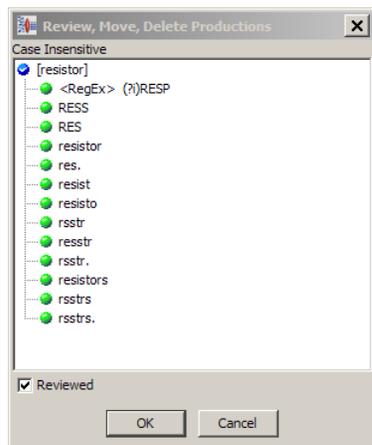
用語バリエーションの生成機能は、ユーザーが設定するデータ・レンズのオプションです。自動生成のバリエーションを作成するデータ・レンズ・オプションを設定すると、用語ルールを作成したときにその生成が適用されます。データ・レンズ・オプションの詳細は、8-2 ページの「[データ・レンズ・オプションの設定](#)」を参照してください。

この機能を有効にする場合は、用語ルールにフルネームのラベルを付ける必要があります。rsstr の例を使用した場合、Knowledge Studio で必要な数のバリエーションを作成するためには、用語ルールに [resistor] というラベルを付ける必要があります。



フレーズ構造の作成を終了すると、バリエーションが定義され、用語のレビューが必要なことを示す、チェック・マーク付きの青色のアイコンが用語の横に表示されます。

自動生成のバリエーションは、用語をダブルクリックすることでレビューできます。



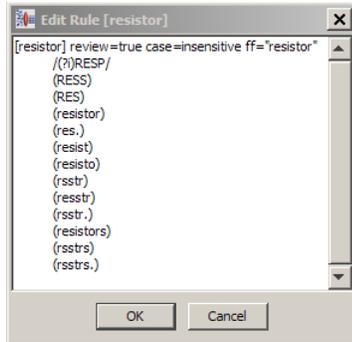
バリエーションをレビューして不要なバリエーションを削除します。リスト内の生成結果をソートするには、任意の生成結果を右クリックし、「**Sort Productions**」を選択します。用語バリエーションのレビューを完了した後は、「**Reviewed**」チェック・ボックスを使用して、レビューの完了をデータ・レンズに示します。

かわりに、用語に「Reviewed」のマーク付けはできませんが、**属性の編集機能**を使用して用語ルールをレビューすることもできます。

## フレーズおよび用語ルールの構文

この項では、データ・レンズのフレーズおよび用語ルールの構文の詳細について説明します。ルールの編集機能を使用すると、ルールの構造を詳細に確認できます。

「Edit Rule」メニュー項目は、「Phrases」タブの「Define Phrases」サブタブからのみ使用でき、用語を右クリックしてアクセスします。



「Edit Rule」ダイアログでは、データ・レンズで現在認識されているルールを変更できます。変更する前に、この機能を使用した場合の結果について十分に理解することをお勧めします。

次の各項では、「Edit Rule」ダイアログを使用してルールの構文を説明します。

### 用語ルール

用語ルールは、大カッコの内側にあるルール名から開始します。この名前に空白は指定できず、文字から開始する必要があります。Knowledge Studio を使用してルールを作成すると、適切な構文が作成されます。次の例では、用語ルール [resistor] に、resistor 用語のバリエーションが含まれています。つまり、[resistor] は、RESS、RESP、RESISTOR などと一致します。カッコ内のテキストは、コンテンツのリテラル・テキストを表しています。各リテラルには、先頭のタブ（空白ではなく）および周囲を囲むカッコが必要です。ルールの各行は生成結果と呼ばれます。各生成結果は、ルールによる認識が可能な一意のテキスト・バリエーションを表しています。

用語ルールを定義した後は、ルールの編集機能を使用して追加のバリエーションをすぐに入力できます。詳細は、10-3 ページの「用語バリエーションの生成」を参照してください。

前述のルール名のすぐ右側にある項目はルール属性です。前述の例の [resistor] ルールでは、大 / 小文字属性が大 / 小文字の区別なしに設定されています。

### 用語ルールにおける大 / 小文字の区別

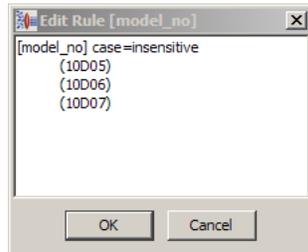
文字列の照合では、大 / 小文字を区別する場合としない場合があります。大 / 小文字の区別は、データ・レンズ・レベルまたは個別の用語ルール・レベルで定義できます。

ただし、正規表現照合の大 / 小文字の区別は、ルール属性として設定できません。それらは、照合コンテキストの観点から設定する必要があります。正規表現定義とその例については、次の項を参照してください。

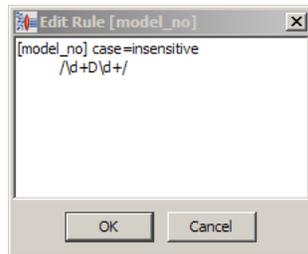
## 用語ルールにおける正規表現

正規表現は、単純な表現で様々なテキスト書式を取得する方法です。たとえば、すべての整数は、正規表現パターンの `/\d+/` で表すことができます。正規表現構文の詳細は、付録 B「正規表現」を参照してください。この項では、標準化プロセスでの正規表現の使用方法について説明します。

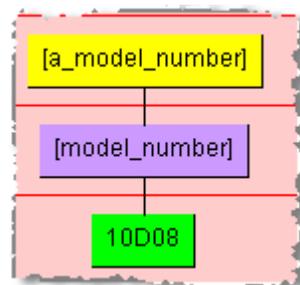
テキストの文字列を含むルールは、文字列と正確に一致する必要があります。これらの文字列は、カッコで直接囲むことによって識別可能です。次に、異なる 3 つのモデル番号を含む用語ルールの例を示します。



コンテンツに表示されているように各エントリを記述するかわりに、モデル番号を整数、D、別の整数の順で定義する、次のショートカットを使用できます。整数の長さは指定されていません。このパターンによって、データ内に出現するこの書式（例：1 つ以上の整数、大文字の D、1 つ以上の整数の順の書式）のすべてのモデル番号が、他の処理なしに照合されます。正規表現は、Knowledge Studio のルールの編集機能を使用して用語ルールに入力できます。



テキスト「10D08」を入力すると、[model\_number] ルールでは、次の例のように認識されます。



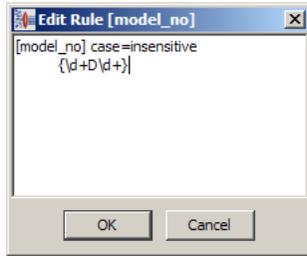
## 空白、正規表現および用語ルール

用語ルール内の正規表現ルールには、2 つの書式があります。

- 空白を区別しない正規表現。これらは、フォワード・スラッシュ (/) 照合文字を使用する用語ルールに定義されます。この正規表現書式は、他の用語に埋め込まれている用語（つまり、空白で囲まれていない用語）を認識する場合に使用できます。
- 空白を区別する正規表現。空白で区切られた正規表現を定義することもできます。通常、この技術は名前と住所に対して使用し、特に埋込みの空白はテキスト品目を識別する際に重要です。これを実行するには、中カッコ ({} ) を使用して正規表現を定義します。あるいは、

10-9 ページの「ソース書式」で説明されているように、ソース書式設定機能を使用して正規表現に空白を追加します。

次の例では、モデル番号を照合しますが、対象は空白で囲まれているモデル番号のみです。



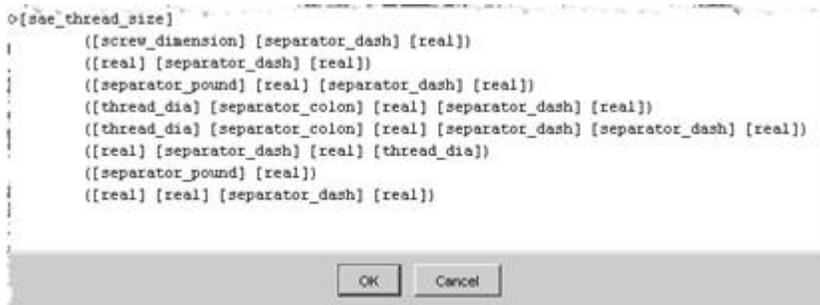
**注意：** 空白区切りの用語を示す中カッコの使用は、Perl 標準に対応した正規表現ではありません。Perl 標準内の中カッコと SCS 定義の中カッコの使用は競合しません。たとえば、次の用語の正規表現は、2つの整数で囲まれた大文字の「D」の後に1から3までの数字が続くモデル番号で、モデル番号全体としては前後に空白があるモデル番号と一致します。

```
{¥d{2}D{1,3}¥d{2}}
```

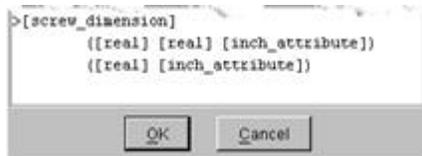
## フレーズ構造ルール

フレーズ構造は、フレーズ・ルールと用語ルールで構成されています。フレーズ・ルールには、用語と他のフレーズで構成された生成結果が含まれています。

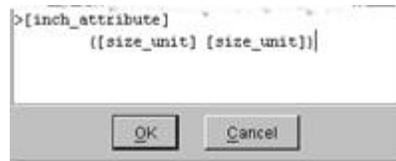
最上位のフレーズ構造ルールは、大なり記号 '>=' から開始する必要があります。次の例では、ルールの各行が [sae\_thread\_size] の様々な書式を表しています。たとえば、最初の書式は、[screw\_dimension] (フレーズ構造自体)、[separator\_dash] および [real] で構成されています。各行はカッコで囲まれ、各フレーズ構造ルール間は空白で区切られています。各フレーズ構造は、1つ以上の用語ルールを参照することになります。他の書式設定は、用語ルールの書式設定と同じです。



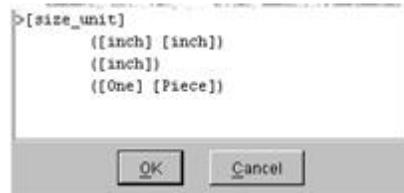
[sae\_thread\_size] は、次のように [screw\_dimension] を参照します。



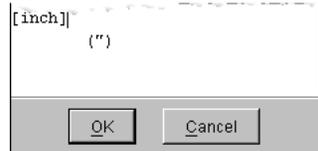
同様に [screw\_dimension] は、次のように [inch\_attribute] を参照します。



同様に [inch\_attribute] は、次のように [size\_unit] を参照します。



最後に、[size\_unit] は、次のように二重引用符でインチを定義する用語ルールの [inch] を参照します。



## フレーズ構造ルールにおける開始記号

フレーズ構造ルールには、開始記号 > と呼ばれる特別な文字を含めることができます。開始記号は、フレーズ構造のルートを示します。

開始記号が含まれるフレーズ構造ルール名は、タグ設定コンテンツ・ファイルにタグとして表示されます。

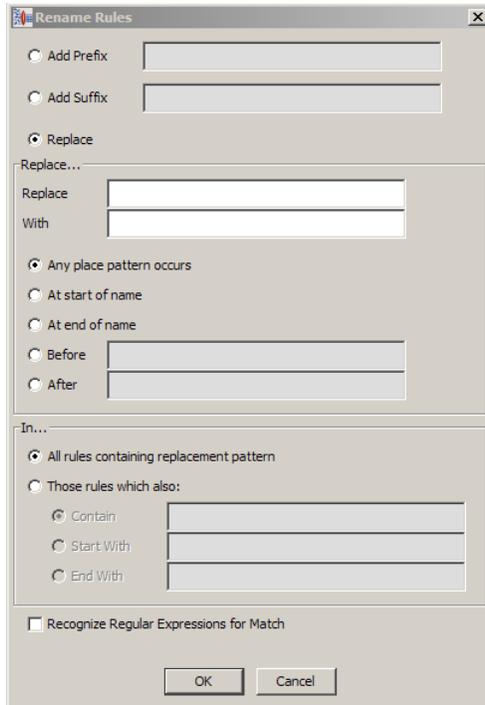
## ルールでの他の制約

ルール名は一意である必要があります。コンテンツの生成結果および属性を表すルール命名規則を使用することをお勧めします。その他の考慮事項：

- フレーズ構造ルールに含めることができるのは、用語ルールの名前のみです。
- 用語ルールに含めることができるのは、リテラル・テキストまたは正規表現のみです。
- ルール名に含めることができるのは、アルファベット、数字およびアンダースコア文字のみです。
- リテラル・テキストの空白には意味があります。(JL) と (JL) は同じではありません。
- ルール名は数字から開始できません。

## グローバル・フレーズ・ルールの名前の変更

フレーズ・ルールの名前を変更し、その変更をデータ・レンズ全体のナレッジに適用する機能は、ルールの一貫性とルールの統合を実現するための強力な方法です。「Edit」メニューから使用できる「Rename Rules」機能には、同じ名前のコンテキスト依存オプションよりも詳細な検索機能が用意されています。



この名前変更操作は、通常の実索および置換に類似しています。接頭辞、接尾辞を追加したり、選択したルールの情報を置換する際に選択できますが、選択できるのは、これらの名前変更機能のうちの1つのみです。

ルールに接頭辞または接尾辞を追加する場合は、単にフィールドにテキストを入力し、「OK」をクリックします。

ルール内の情報の置換では、データの検索および置換に関する自明の選択肢が提供されます。「Replace」フィールドと「With」フィールドの両方を空白のままにすることはできませんが、いずれかが空白の場合は、選択した他のオプションに基づいて削除として機能します。

たとえば、「At start of name」 / 「At end of name」、「Before」および「After」オプションを使用して、変更するルール内の場所を指定する場合があります。オプションの使用方法は次のとおりです。

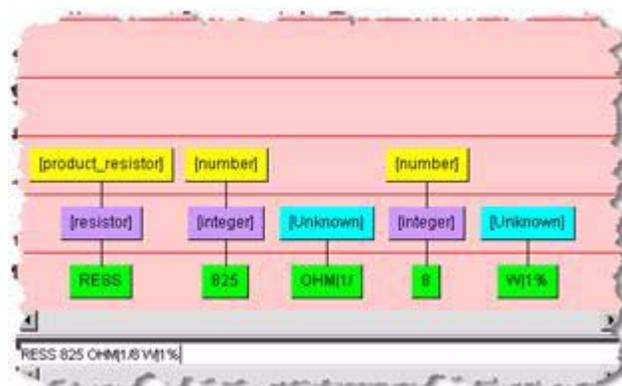
<b>Replace</b>	res_
<b>With</b>	resistor_
<b>After</b>	a_

前の例では、「a\_res\_ohms」が「a\_resistor\_ohms」に変更されますが、「a\_resolution\_pixels」は変更されません。

## ソース書式

ソース書式を使用すると、ナレッジを作成する前にコンテンツの書式を再設定できます。書式設定の目的は、コンテンツの標準化に対する特殊ルールを減らすことです。データ・レンズ・ルールの適用前に実行する必要があるのは、ソース書式設定のみです。

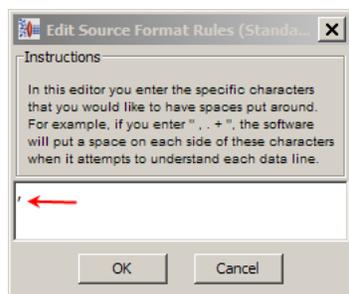
たとえば、次の図は、「**Graphical Rule Building**」ペイン直下の「**Selected Content**」フィールドに入力したテキストの「Res10Ohm|1¥8W|」を示しています。この例では、データにパイプ (|) フィールド・セパレータが含まれています。Knowledge Studio では、用語を「|」で区切ることはできません。このような問題がサンプル・データ内で繰り返し検出される場合は、ソース書式設定を適用する必要があります。



## ソース書式設定の適用

1. 「Data Lens」メニューから「Source Format」をクリックして、「Standard Mode」を選択します。
2. 「Edit Source Format Rules」ダイアログの入力フィールドにカンマを入力し、「OK」をクリックします。

たとえば、データ内のすべてのカンマの周囲に空白を追加する必要がある場合は、「,」（カンマ）を入力します。



3. 「OK」をクリックします。
4. テキスト行に新しいソース書式設定を適用するには、「Sample Data」ペイン内をクリックし、[Enter] キーを押します。
5. 「Refresh」をクリックしてソース書式設定をリフレッシュし、サンプル・データ・ファイル全体を変更します。

「Graphical Rule Builder」ペインでは、次のようにフレーズ構造表示が更新されます。



---

## データの詳細な標準化

この章では、Knowledge Studio を使用した製品データの標準化に関連する技術および情報について説明します。

## 用語の標準化

次の各項では、ルールを使用して用語を詳細に標準化する方法について説明します。

### 大 / 小文字の置換

文字列の大 / 小文字の置換を上書きする機能があります。たとえば、すべての文字列置換が小文字に設定されていて、すべてを大文字に切り替える必要がある場合は、文字列置換タブの「Case」オプションを選択して簡単に上書きできます。

たとえば、「resistor」という用語が最初に先頭大文字の「Resistor」に置換された場合、用語のリライト・ルールの「Case」オプションを使用して、用語をすべて大文字のデータに置換できます。

**Rewrite rule for [resistor]**

No Replacement

Replace All

Resistor

Regular Expression Replacement

Individual Replacement

Original	
RESS	
RESP	
RES	

Add

Case

Set Case

TO UPPERCASE

to lowercase

To Proper Case

この設定変更の結果として、「resistor」という用語のすべてのインスタンスが大文字に変更されます。

Initial Text	Standardized Text
RESPCF AX300OHM1/4W5%	RESISTOR 300 Ohm 5% 0.25 Watt Axial Carbon Film
220 ohm CF AX 0.5W 5 % resp	RESISTOR 220 Ohm 5% 0.5 Watt Axial Carbon Film
array 16 pin 85 ohm 5% resp	RESISTOR 85 Ohm 5% 16 Pin Array
NTWK 330 OHM 8 P 2% RESP	RESISTOR 330 Ohm 2% 8 Pin Network
CF, AX 75 OHM 1/4W 5% RESP	RESISTOR 75 Ohm 5% 0.25 Watt Axial Carbon Film

## 正規表現の置換

正規表現の置換は、置換対象テキストが変数の可能性がある場合の複雑な文字列置換を意図しています。この項の例では、2桁の年から4桁の年への高度な変換を処理し、事前に定義された境界に基づいて正しい世紀の年を設定する方法を示します。

2桁の年に対するルールが定義されており、2桁の年から4桁の年に変換する必要があると仮定します。正しい世紀で年を設定するには、29以下の年をすべて21世紀に設定し、30以上の2桁の年をすべて20世紀に設定します。

最初に、次のように正規表現を使用して年に対する用語ルールを作成します。

```
[year] case-insensitive
/[0-2][0-9]/
/[3-9][0-9]/
```

次に、2桁から4桁の年を作成するために、次のような正規表現文字列の置換を作成します。

**Rewrite rule for [year]**

No Replacement

Replace All

Regular Expression Replacement

s/[0-2][0-9]/20\$1/;s/[3-9][0-9]/19\$1/

**注意：** この例の文字列は、必要な結果を得るために、複数の正規表現の置換が併記されています。各表現をセミコロン (;) で区切ることで、単一の正規表現置換として使用できます。

この例の文字列の置換は、結果的に2桁の年が次のように標準化されます。

Source Text (editable - press Enter when done)

10 20 30 40 50 60

Formatted Text

10 20 30 40 50 60

Standardized Text

2010 2020 1930 1940 1950 1960

予想のとおり、30未満の2桁の年にはすべて前に数字の20が付き、30以上の2桁の年にはすべて前に数字の19が付きます。

**注意：** 正規表現をよく理解していない場合は、コンサルタントの指導を受けた後で前述の置換を使用してください。

**注意：** 詳細は、B-1ページの「正規表現」を参照してください。

## 個別の置換

用語を個別に置換する場合、用語ルールの大 / 小文字の区別は置換ルールから継承されます。たとえば、用語ルールが大 / 小文字の区別なし（デフォルト）に定義されている場合、置換ルールは次のようになります。

```
[resistor] case=insensitive
  (RESS)
  (res)
  (res.)
  (resistor)
  (resistors)
  (rstr)
  (rstr.)
```

Individual Replacement	
Original	Rewrite As
RESS	
res	Resistor
res.	
resistor	
resistors	
rstr	
rstr.	

前述の置換ルールでは、「RES」と「res」の両方を「Resistor」というテキストに置換します。

## 複数の標準化タイプ

標準化タイプは、「**Standardization Types**」ダイアログを使用して必要に応じていくつでも構成できます。メリットは、単一のデータ・レンズ・プロジェクトに定義されているフレーズと用語を再利用して任意の数の標準結果を定義できることです。

たとえば、Web サイトで使用するために、略語を使用せずに理解しやすい長い形式の摘要を作成する標準化タイプが必要です。また、データベース表に起因する文字の長さ制限に適合するように、短縮された摘要を作成する短縮形の標準化タイプも必要になります。

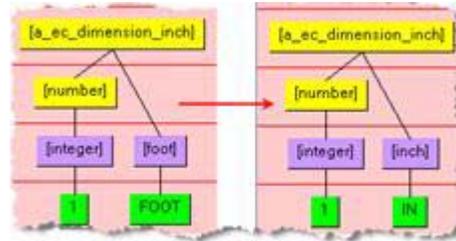
詳細は、3-12 ページの「[標準化タイプの作成](#)」を参照してください。

## 単位の標準化タイプ

標準化の要件としてデータを単一の単位に換算する必要がある場合は、単位換算タイプを使用することで処理できます。

用語およびフレーズのルールに対して複数の単位が定義されたデータが存在する場合は、異なる単位を使用するようにこれらのルールを定義し、標準の摘要を提供するために単一の単位に換算できます。

たとえば、[a\_ec\_dimension\_inch] のデータをインチまたはフィートにできます。データを標準化する場合、必須の単位はインチです。次のようにフィートからインチに換算するフレーズに対してテキストが解析されます。

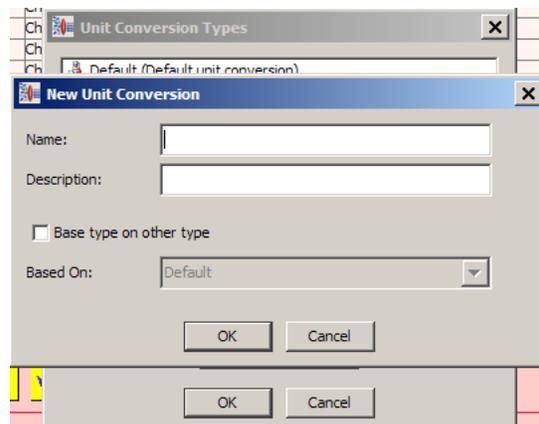


データ・レンズ全体で使用する独自の標準化スキーマを作成できます。標準化タイプは、多様な用途に使用できます。

**注意：** 各データ・レンズを使用して Knowledge Studio で作成されたデフォルトの整数、分数、小数のルールを変更すると、単位換算が失敗します。たとえば、「one」というテキストを [integer] ルールに追加する場合などです。

## 単位換算タイプの作成

1. 「Data Lens」メニューから「Unit Conversion Types...」を選択します。
2. 「Add New」ボタンをクリックします。



3. 新しい単位換算タイプの作成に必要な情報を入力し、選択オプションとして「Unit Conversion Types」リストに追加します。
4. 作成した単位換算タイプがすでにあり、同じ標準化の新しいバージョンでそのナレッジを再利用する場合は、「Base type on other type」チェック・ボックスを選択し、次に「Based On:」リストから適切な分類タイプ選択します。

---

---

**注意：** 他の単位換算タイプがない場合、このチェック・ボックスはアクティブになりません。

---

---

5. 「OK」をクリックします。

標準化タイプと同様に、データ・レンズには複数の単位換算タイプを関連付けることができます。

## 単位換算タイプの削除

単位換算タイプは、必要に応じて削除できます。

1. 最新のデータ・レンズ・バージョンをチェックインしたことを確認します。
2. 「Data Lens」メニューから「Unit Conversion Types...」を選択します。
3. 削除する単位換算タイプを選択して右クリックします。
4. 「Delete Unit Conversion」をクリックします。  
削除確認ダイアログが表示されます。
5. 選択した単位換算タイプを削除する場合は「OK」をクリックし、削除しない場合は「Cancel」をクリックします。
6. 「OK」をクリックします。

## フレーズに対する単位換算の作成

フレーズの単位換算を作成する場合は、単位換算する生成結果ごとに数字と単位を設定する必要があります。設定しないと換算が失敗します。

1. 「Standardize」タブおよび「Unit Conversion」サブタブを選択します。
2. 単位換算が必要なフレーズの生成結果を選択します。

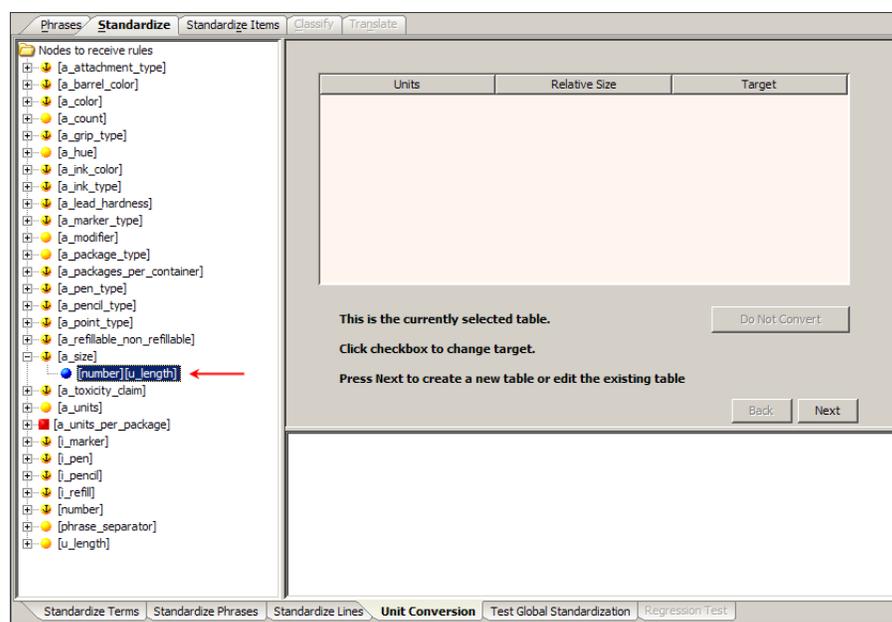
---

---

**注意：** 単位の標準化タイプが関連付けられていないフレーズには、フレーズの横に青色の丸いアイコンが付きます。単位換算があるフレーズには、紫色の丸いアイコンが付きます。換算された生成結果の親のフレーズは、赤色の四角形から黄色の丸いアイコンに変更されます。

---

---



3. 「Next」 をクリックして、単位換算表の作成を開始します。

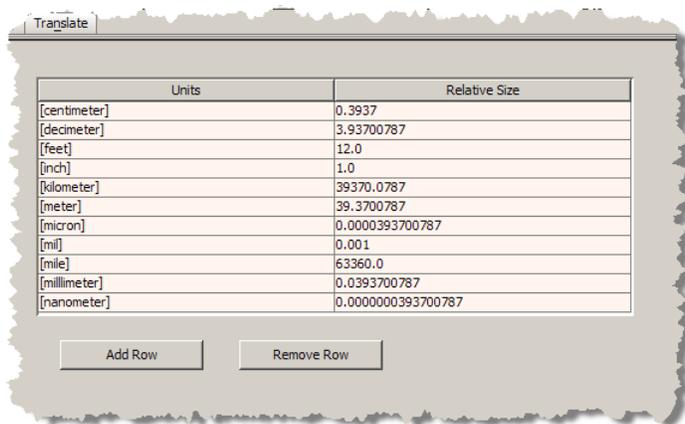


4. 「Next」 をクリックし、換算について、「Number」に [number]、「Unit」に [u\_length] というルールを選択を受け入れます。
5. 既存の表名を選択するか、新しい表名を入力します。

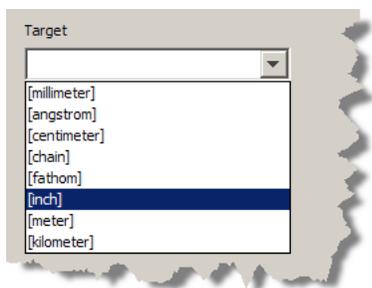


6. この情報が正しい場合は、「Next」 をクリックします。正しくない場合は、このフレーズのドロップダウン・ボックスから使用する正しい表を選択するか、新しい表を作成して選択します。
7. 「Next」 をクリックしてウィザードを進めます。

既存の単位換算がある表を選択した場合は、単位換算が表に表示されます。単位換算がない場合は、空白の表が表示されます。



8. 新しい表および新しい単位換算が作成された場合は、「Add Row」および「Remove Row」ボタンを使用して単位表を作成する必要があります。
9. ダイアログが表示された後、単位名を入力し、対応するサイズ・フィールドに換算係数を指定します。
10. 必要に応じて行をすべて完成し、「Next」をクリックします。



11. ターゲット単位を選択します。  
たとえば、フィートが含まれているデータをインチに換算する必要がある場合は、[inch]というターゲットを選択してください。
12. 「Next」をクリックします。

フレーズの横のアイコンが、単位の標準化タイプがフレーズに関連付けられたことを示す紫色に変わり、関連するフレーズ・ルールの前に赤色の四角形が表示されます。

単位の標準化をテストするには、「Test Standardization」サブタブを選択し、換算対象テキストを入力します。

#	...	Initial Text	Standardized Text
1	100 34	010MF 10V 10 SMT 0603 XZR	0.010 10 16 SURFACE MOUNT XZR 0603
2	100 34	010MF 16V 10 SMT 0603 XZR	0.010 10 16 SURFACE MOUNT XZR 0603
3	100 32	01MF 50V 5 SMT 1206 XZR	0.01 5 50 SURFACE MOUNT XZR 1206
4	100 32	01MF 50V 5 SMT 1206 XZR	0.01 5 50 SURFACE MOUNT XZR 1206
5	100 34	030MF 10V 10 SMT 0603 XZR	0.033 10 10 SURFACE MOUNT XZR 0603
6	100 34	033MF 10V 10 SMT 0603 XZR	0.033 10 10 SURFACE MOUNT XZR 0603
7	100 32	1MF 16V 10 SMT 0603 XZR	0.1 10 16 SURFACE MOUNT XZR 0603
8	100 32	1MF 16V 10 SMT 0603 XZR	0.1 10 16 SURFACE MOUNT XZR 0603
9	100 31	1MF 50V 5 SMT 1206 XZR	0.1 5 50 SURFACE MOUNT XZR 1206
10	100 31	1MF 50V 5 SMT 1206 XZR	0.1 5 50 SURFACE MOUNT XZR 1206
11	100 33	22MF 10V 10 SMT 0603 XZR	0.22 10 10 SURFACE MOUNT XZR 0603
12	100 33	22MF 10V 10 SMT 0603 XZR	0.22 10 10 SURFACE MOUNT XZR 0603
13	100 37	10MF 50V 20 SMT 6032 TNT	1 0 20 50 SURFACE MOUNT TANTALUM 6032
14	100 37	10MF 50V 20 SMT 6032 TNT	1 0 20 50 SURFACE MOUNT TANTALUM 6032
15	100 36	10MF 20V 20 SMT 6032 TNT	10 20 20 SURFACE MOUNT TANTALUM 6032
16	100 36	10MF 20V 20 SMT 6032 TNT	10 20 20 SURFACE MOUNT TANTALUM 6032
17	100 37	3.3MF 35V 20 SMT 6032 TNT	3.3 20 35 SURFACE MOUNT TANTALUM 6032
18	62 47	3.3MF 35V 20 SMT 6032 TNT	3.3 20 35 *** SURFACE MOUNT * TANTALUM * 6032

単位換算によって、入力したテキストが正しい標準化単位に変換されます。

単位換算をオンにすると標準化属性の下にある範囲が使用可能になり、さらに値および検索ロジックを品目定義で使用できるようになります。これらのメリットを活用するには、単位換算の選択の必要がなくても単位換算を作成する必要があります。たとえば、分数を出力する場合は、値ロジックと範囲が引き続き正しく機能するように、単位換算を無効にせず、単位換算をなしに設定してください。

## データ・レンズでの標準化の共有

標準化ナレッジを作成する取組みはデータによっては相当な労力になるため、様々な標準化タイプや品目定義に情報を繰り返し入力しなくてすむように、標準化ルールを共有（再利用）する機能を使用して人件費を節約します。既存の標準化タイプに基づいて新規のタイプを作成する機能が有用であることに加え、標準化ナレッジを調整した後で共有することで、データを詳細に定義する次のステップが簡単になります。

次の各項では、ルールのコピー方法について説明します。

- 標準化タイプ間でのグローバル標準化ルールのコピー
- 別の標準化タイプへの品目定義の標準化ルールのコピー
- 品目定義間での標準化ルールのコピー

これらの項の全体を通して、次の用語が使用されます。

### ソース

コピーするナレッジ（標準化ルール）が含まれている標準化タイプまたは品目定義。

### ターゲット

選択した一連のナレッジを受け取るコピー先の標準化タイプまたは品目定義。

### スコープ

ナレッジをソースからターゲットにコピーする方法（新規のみ、マージまたは置換）。各コピー操作に伴って Knowledge Studio により、最上位レベル（フレーズまたは親品目定義）から最下位レベル（生成結果または属性）まで、ソースとターゲットの両方が点検され、各ルール、生成結果および表（該当する場合）の相違が判断されます。

### コンテンツ

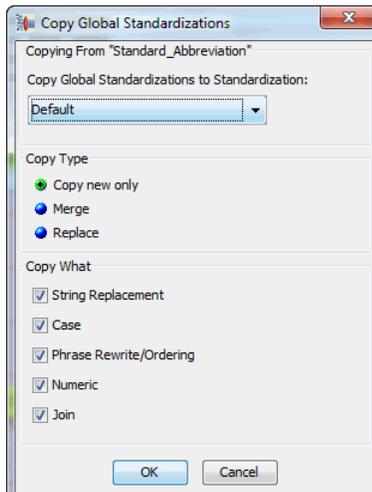
リストで選択したソースからターゲットにコピーする標準化ルール。たとえば、ある標準化タイプの大 / 小文字設定ルールを変更して、データ・レンズ内のその他の標準化タイプにコピーすることで、この変更がすぐに反映され、当初は小文字に標準化されていたデータを簡単に大文字に変更できます。

## グローバル標準化のコピー

ある標準化タイプで調整したグローバル標準化ルールをデータ・レンズ内の別の標準化ルールにコピーできます。

現在の標準化タイプがソースであり、標準化ルールのコピー先のターゲット標準化タイプを選択するため、このオプションは Knowledge Studio のあらゆるタブで使用できます。

1. ツールバーの「Standardization Type」リストから、ソース標準化タイプを選択します。
2. 「Data Lens」メニューから「Copy Global Standardizations」をクリックします。



3. リストからターゲット標準化タイプを選択します。
4. 次のいずれかのコピー・スコープ・オプションを選択して、選択した標準化タイプからグローバル標準化ルールをコピーします。

- **Copy new only**

ソース標準化タイプに存在し、ターゲット標準化タイプに存在しないソース標準化タイプのコンテンツ（次のステップで選択）の標準化ルールのみがコピーされます。つまり、ターゲット標準化タイプにとって新しい標準化がコピーされます。

- **Merge**

ソース標準化タイプのコンテンツ（次のステップで選択）の標準化ルールが、ターゲット標準化タイプにマージされます。ターゲット標準化タイプにない標準化ルールのみがコピーされます。たとえば、リライト表の新規の生成結果とエントリがコピーされます。競合が検出された場合、そのルールは無視され、コピーされません。

- **Replace**

ソース標準化タイプのコンテンツ（次のステップで選択）のすべての標準化が、既存の標準化ルールを上書きしてターゲット標準化タイプにコピーされます。

空白のソース標準化タイプ（ルールが格納されていない）をコピーし、ルールが格納されているターゲットを上書きしようとしていることが、Knowledge Studio で検出された場合は、その操作が不可能であることを示すメッセージが表示され、コピーが終了します。ソースおよびターゲットの標準化タイプをレビューし、指定が正しいことを確認してください。

5. 次のチェック・ボックスを使用して、標準化タイプとともにコピーするコンテンツを選択します。

- **String Replacement**

「Standardize」タブの「Standardize Terms」サブタブに表示されるすべてのテキスト置換表（「No Replacement」、「Replace All」、「Regular Expressions Replacement」、

「Individual Replacement」の表を含む) リライト・ルール of 個々の文字列置換ルールすべて。

- **Case**

「Standardize」タブの「Standardize Terms」サブタブに表示される個々の大 / 小文字置換ルールすべて。標準化タイプのデフォルトの大 / 小文字設定は含まれません。

- **Phrase Rewrite/Ordering**

「Standardize」タブの「Standardize Phrases」サブタブに定義されている用語およびフレーズの順序ルールすべて。

- **Numeric**

テキストと換算表がある「Standardize」タブの「Standardize Terms」サブタブに定義されている、すべての値範囲リライト・ルール表の数値テキスト置換ルールすべて。

- **Join**

「Standardize」タブの「Standardize Phrases」サブタブに定義されている（生成結果レベルではなく）フレーズ・レベルの用語結合ルールすべて。

## 6. 「OK」をクリックします。

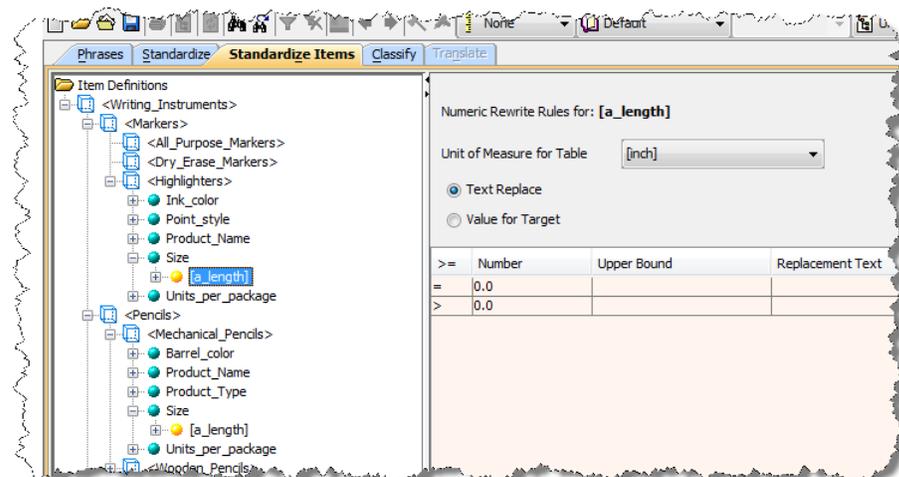
選択したスコープおよびコンテンツを使用して、グローバル標準化ルールがソースからターゲットの標準化タイプにコピーされます。完了時に、変更の詳細を示す確認メッセージが表示されます。

ターゲット標準化タイプを選択してレビューし、ルールが正しくコピーされたことを確認してください。

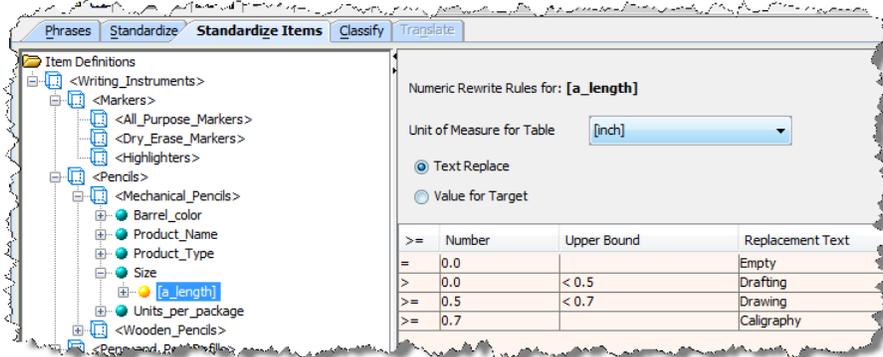
## 品目定義標準化の共有

品目定義標準化ルールは、単一の標準化タイプ内および複数の標準化タイプにわたって、他の品目定義と共有できます。通常は、品目定義を作成して標準化した後で共有します。この機能では、継続的に品目定義を修正し、標準化ルールを他の品目定義および標準化タイプにコピーすることもできます。

この機能は、ソースとターゲットの両方の品目定義に存在する同じ属性、フレーズおよび生成結果構造に依存します。品目定義間で標準化ルールをコピーしても、(ソースに存在する) 属性はターゲットに作成されません。次の例では、Highlighters と Mechanical\_Pencils の両方の品目定義に、[a\_length] という生成結果が設定された Size 属性があります。Size と [a\_length] の標準化ルールは、任意の品目定義から別の品目定義にコピーできます。



共有できるのは、両方の品目定義に存在する属性のフレーズの生成結果に対する標準化ルールです。次の例では、Mechanical\_Pencils の [a\_length] という生成結果に対する標準化ルールを Highlighters と共有できます。



ある品目定義から別の品目定義に標準化ルールをコピーするための3つのオプションがあります。

#### Copy to...

現在選択されている標準化タイプのソース品目定義からターゲット標準化タイプのターゲット品目定義に、標準化ルールをコピーします。子品目定義が含まれている場合でも、親品目定義のみがコピーされます。

#### Copy with Children to...

現在選択されている標準化タイプの親品目定義とすべての子品目定義から、ターゲット標準化タイプの品目定義の同じセットに標準化ルールをコピーします。

#### Copy to another Item Definition

現在選択されている標準化タイプ内で、ソース品目定義からターゲット品目定義に標準化ルールをコピーします。両方の品目定義に共通する属性、フレーズおよび用語ルールの標準化ルールのみがコピーされます。

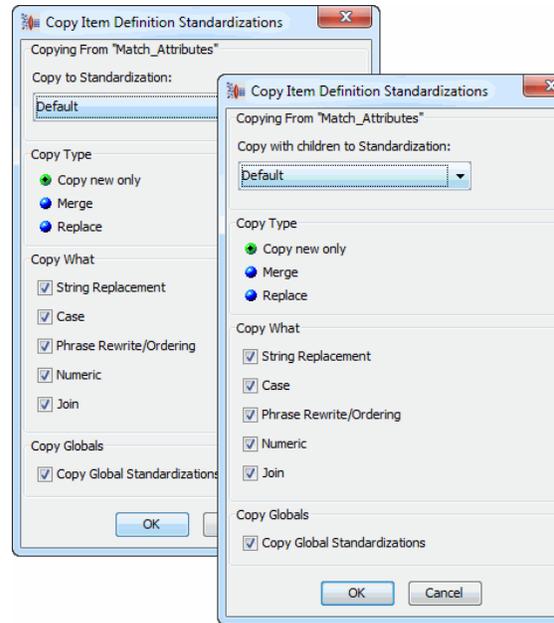
コピーする標準化ルールのスコープと深さは、ユーザーが選択します。

**ヒント：** コピーできるのは、ターゲット標準化タイプまたは品目定義の非アクティブな品目定義とアクティブな品目定義です。

### 標準化タイプ間での品目定義のコピー

1. ツールバーの「Standardization Type」リストから、ソース標準化タイプを選択します。
2. 「Standardize Items」タブおよび「Standardize Attributes」サブタブを選択します。
3. コピーする品目定義を右クリックし、「Copy to...」または「Copy with Children to...」オプションを選択します。

**注意：** 子がない品目定義を選択すると、「Copy with Children to...」オプションは使用できません。このオプションはコンテキスト依存です。



ソース品目定義の共有範囲が前述の範囲と異なるとしても、これらの2つのダイアログの使用方法は同じです。

4. 標準化タイプのリストから、ソース品目定義のコピー先となるターゲット標準化タイプを選択します。
5. 次のいずれかのオプションを選択して、ソース品目定義のコピー方法を定義します。

- **Copy new only**

ソースに存在し、ターゲット標準化タイプのターゲット品目定義に存在しないコンテンツ（次のステップで選択）の標準化ルールのみがコピーされます。つまり、ターゲット標準化タイプのターゲット品目定義にとって新しい標準化ルールがコピーされます。

- **Merge**

ソース標準化タイプのソース品目定義のコンテンツ（次のステップで選択）の標準化ルールが、ターゲット標準化タイプのターゲット品目定義にマージされます。ターゲットにない標準化ルールのみがコピーされます。競合が検出された場合、そのルールは無視され、コピーされません。

- **Replace**

ソース標準化タイプのソース品目定義のコンテンツ（次のステップで選択）の標準化ルールすべてが、既存の標準化ルールを上書きしてターゲット標準化タイプのターゲット品目定義にコピーされます。

---

**注意：**「Standardize Items」タブの「Order Attributes」サブタブおよび「Match Weights」サブタブで定義した、品目定義の属性の順序と照合ランク付けの標準化は、品目定義をコピーする際のコピー対象には含まれません。

---

6. 次のチェック・ボックスを使用して、品目定義とともにコピーするスコープを選択します。
  - **String Replacement**  
 「Standardize Items」タブの「Standardize Attributes」サブタブに表示されるすべてのテキスト置換表（「No Replacement」、「Replace All」、「Regular Expression Replacement」、「Individual Replacement」の表を含む）リライト・ルール個々の文字列置換ルールすべて。
  - **Case**  
 「Standardize Items」タブの「Standardize Attributes」サブタブに表示される大 / 小文字置換ルールすべて。
  - **Phrase Rewrite/Ordering**  
 「Standardize Items」タブの「Standardize Attributes」サブタブに定義されている用語およびフレーズの順序ルールすべて。
  - **Numeric**  
 テキストと換算表がある「Standardize Items」タブの「Standardize Attributes」サブタブに定義されている、すべての値範囲リライト・ルール表の数値テキスト置換ルールすべて。
  - **Join**  
 「Standardize」タブの「Standardize Phrases」サブタブに定義されている（生成結果レベルではなく）親フレーズ・レベルの用語結合ルールすべて。  
  
 このチェック・ボックスは、「Copy Global Standardizations」チェック・ボックスが選択されている場合のみアクティブになります。
7. 「Copy Global Standardizations」チェック・ボックスを選択し、スコープおよびコンテンツの選択に基づいて、ソース品目定義内でグローバル標準化ルールをコピーします。このオプションはスティッキーで、このダイアログを使用するたびに前回の選択が再表示されます。

---

**注意：** このオプションを選択し、標準化のコピーで変更が発生しなかった場合は、変更のグローバル・コピーが終了し、コピーは失敗となります。これは、品目定義レベルで変更がなかった場合、グローバル変更のコピーにはリスクが伴うためです。品目定義標準化ルールとともにコピーを意図したものをレビューしてください。グローバル標準化を強制的にコピーする方法は、11-10 ページの「[グローバル標準化のコピー](#)」を参照してください。

---

8. 「OK」をクリックします。

選択したスコープおよびコンテンツを使用して、ソース標準化タイプの品目定義標準化ルールがターゲット標準化タイプのターゲット品目定義にコピーされます。完了時に、変更の詳細を示す確認メッセージが表示されます。

ターゲット標準化タイプおよび品目定義を選択し、標準化ルールをレビューします。必要な結果が得られたことを確認します。

---

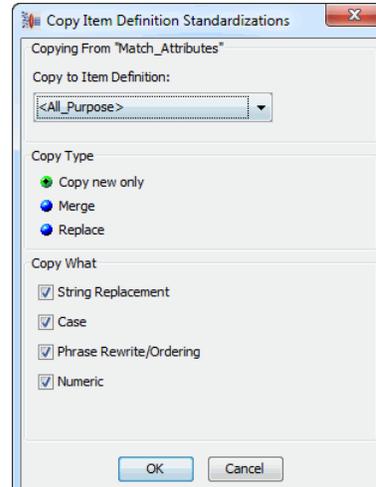
**注意：** 「Replace」オプションを使用する場合、ターゲット標準化タイプで、ソース標準化ルールの 20% を超えて置換しようとする、警告が表示されます。これは、ターゲットの既存のルールが上書きされ、有用な標準化ナレッジが消失するのを回避することを意図しています。

---

保持しておく必要がある標準化ルールが不用意に上書きされる可能性もあるため、ソースおよびターゲットのコピー結果は慎重にレビューしてください。

## 品目定義間のコピー

1. ツールバーの「Standardization Type」リストから、ソース標準化タイプを選択します。
2. 「Standardize Items」タブおよび「Standardize Attributes」サブタブを選択します。
3. コピー元のソース品目定義を右クリックし、「Copy to another Item Definition」オプションを選択します。



4. ソース品目定義をコピーするターゲット品目定義をリストから選択します。
5. 次のいずれかのオプションを選択して、品目定義のコピー方法を定義します。

- **Copy new only**

ソースに存在し、ターゲット品目定義に存在しないコンテンツ（次のステップで選択）の標準化ルールのみがコピーされます。つまり、属性の既存のフレーズの生成結果に関する標準化ルールの中で、ターゲット品目定義にとって新しいルールがコピーされます。

- **Merge**

ソース品目定義のコンテンツ（次のステップで選択）の標準化ルールが、ターゲット品目定義にマージされます。ターゲットにない標準化ルールのみがコピーされます。競合が検出された場合、そのルールは無視され、コピーされません。

- **Replace**

ソース品目定義のコンテンツ（次のステップで選択）のすべての標準化ルールが、既存の標準化ルールを上書きしてターゲット品目定義にコピーされます。

---

**注意：**「Standardize Items」タブの「Order Attributes」サブタブおよび「Match Weights」サブタブで定義した、品目定義の属性の順序と照合ランク付けの標準化は、品目定義をコピーする際のコピー対象には含まれません。

---

6. 次のチェック・ボックスを使用して、「**Standardize Items**」タブの「**Standardize Attributes**」サブタブに定義されている品目定義とともにコピーするスコープを選択します。
  - **String Replacement**

すべてのテキスト置換表（「No Replacement」、「Replace All」、「Regular Expressions Replacement」、「Individual Replacement」の表を含む）リライト・ルール個々の文字列置換ルールすべて。
  - **Case**

大 / 小文字置換ルールすべて。
  - **Phrase Rewrite/Ordering**

用語およびフレーズの順序ルールすべて。
  - **Numeric**

テキストと換算表が含まれているすべての値範囲リライト・ルール表の数値テキスト置換ルールすべて。
7. 「OK」をクリックします。

選択したスコープおよびコンテンツを使用して、ソース品目定義ルールがターゲット品目定義にコピーされます。完了時に、変更の詳細を示す確認メッセージが表示されます。「Item Definition」ペインがリフレッシュされ、展開されていた品目定義が縮小されます。

ターゲット品目定義を選択してレビューし、正しくコピーされたことを確認してください。

---

## データの詳細な分類

この章では、Knowledge Studio を使用したデータの分類に関連する技術および情報について説明します。

## 高度な分類ルールの機能

この項では、データの分類の絞込みに使用できる様々な分類機能について例を含めて説明します。

### 追加

追加機能は、2つ以上の文法を指定し、その結合により品目の分類を定義することを意図しています。

たとえば、「Power Nail Stapler」と「Paper Stapler」について考えてみます。Stapler は事務用品の Stapler を分類するには十分ですが、Power Nail Stapler を電動工具として正しく分類するには追加の属性が必要です。そのため、Stapler という品目に Nail を追加します。

### マスキング

マスキング機能は、マスクされたフレーズの下にあるすべての文法を無視することを意図しています。一般的には、分類するプライマリ品目の一部でないものが含まれている場合に使用されます。

たとえば、「Drill with Charger」について考えてみます。ここで、品目は Drill であり、Charger ではありません。次の例に、マスキングの使用と関連する文法を示します。

### 否定

否定機能は、包含、または明記されていない前置詞が示唆される文法すべてを無視することを意図しています。

たとえば、「Toner Cartridge HP Printer」について考えてみます。品目は Toner Cartridge であり、プリンタではありません。

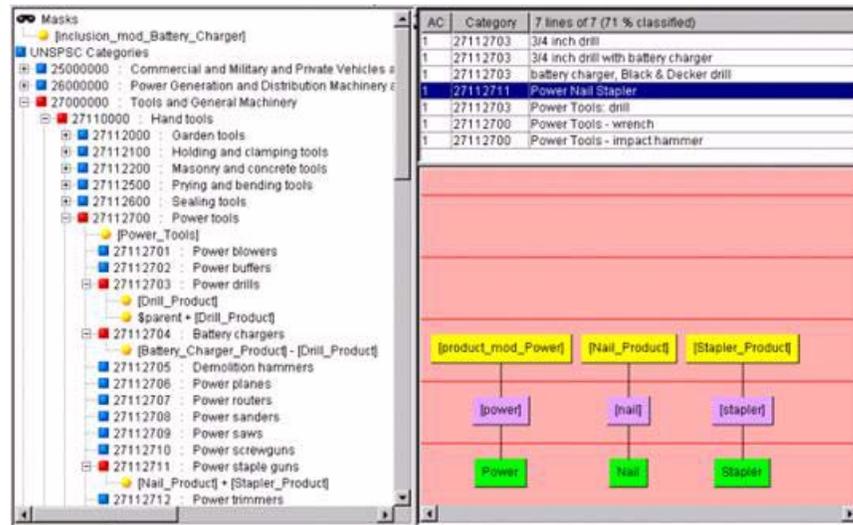
### 親

親機能は、分類ツリーの上位レベルで文法を参照することを意図しています。その用途は、上位レベルから他の特異な属性が定義されている下位レベルまで継承を適用することです。

たとえば、レジスタには固定と変動の両方のタイプがあります。ツリーで [product\_resistor] という用語はスキーマのレジスタ・レベルに存在し、変動や固定はレジスタ・レベルより下位のレベルに存在します。[product\_resistor] と [attr\_variable] 間の接続は、\$parent + [attr\_variable] (\$parent は [product\_resistor] を参照する) という用語を介して実行されます。データを上位レベルで一括分類して初期の分類を取得し、その後の段階で分類を絞り込むことが有用です。

## 機能の例

この例は、分類の複雑化を引き起こす品目の集まりを示しています。前述の機能を使用することで、あいまい表現が削除され、各品目を一意に分類できるようになります。



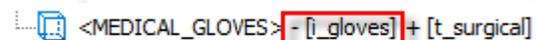
### マスクの追加

マスキングを追加するには、マスクするルート・レベルの文法を、分類ツリーの上部のマスク・アイコン上にドラッグします。

この文法の下に表示されているすべての文法は、さらに分類されないように非表示になります。ただし、マスクされた文法の下で非表示になっている文法は、分類に使用できる他のフレーズ構造には表示されます。

### 否定の追加

否定を追加するには、[Ctrl] キーと [Shift] キーを同時に押したまま、関連するプライマリ分類品目の横に否定フレーズをドラッグします。

 <MEDICAL\_GLOVES> - [i\_gloves] + [t\_surgical]

### 親の追加

親を追加するには、次の手順を使用します。

1. プライマリ品目を上位レベル分類ノードにドラッグします。
2. セカンダリ品目を下位レベル分類ノードにドラッグします。
3. セカンダリ品目を右クリックし、「Add Parent」を選択します。

## 複数の分類スキーマ

分類のタイプは、分類タイプ機能を使用して必要に応じていくつでも構成できます。メリットは、単一のデータ・レンズ・プロジェクトに定義されているフレーズと用語を再利用して任意の数の分類結果を定義できることです。

たとえば、UNSPSC スキーマにデータを分類すると同時に、eCl@ss スキーマやユーザー定義のスキーマにもデータを分類できます。

新しい分類タイプを作成するプロセス。詳細は、5-3 ページの「[分類タイプ](#)」を参照してください。

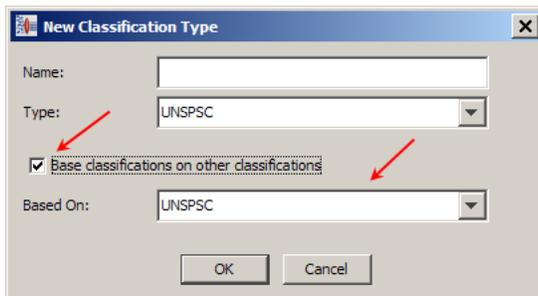
複数の分類スキーマを構成する場合は、次の点を考慮する必要があります。

- 新しい分類タイプ名を作成する場合は、名前の一部に分類バージョン番号情報を使用して区別できるようにしてください。たとえば、UNSPSC 11.1 を使用している場合は、UNSPSC\_11\_1 などの名前を使用します。
- 前の分類タイプにすでに作成されているルールを再利用するには、「Base classifications on other classifications」チェック・ボックスを選択し、新しいタイプのベースにする分類タイプを選択します。

## 分類タイプのアップグレード

ある時期に、特定の分類タイプを同じ分類タイプのさらに新しいバージョンにアップグレードすることが必要な場合があります。たとえば、UNSPSC 分類バージョン 9.2 をさらに新しいバージョンの 11.1 にアップグレードできます。新しい分類タイプにアップグレードし、前のバージョンのナレッジはすべて保持できます。

ある分類タイプから新しいタイプにアップグレードするには、新しいタイプが既存のタイプに基づいている必要があります。新しい分類タイプの作成プロセスを使用する場合は、「Based On:」リストから新しいタイプのベースにする適切なスキーマを選択できるように、必ず「Base classifications on other classifications」チェック・ボックスを選択します。詳細は、5-3 ページの「[分類タイプ](#)」を参照してください。



新しいバージョンに存在せず、前のバージョンには存在する分類マッピングが、Knowledge Studio で検出された場合は、カテゴリ構造の変更の性質を示すメッセージが表示されます。

両方の分類スキーマがロードされ、黒色の矢印を使用して 2 つを切り替えることができます。

## ユーザー定義の分類タイプ

Knowledge Studio では、自動分類に使用できる独自のカスタム・スキーマを作成できます。これはユーザー定義の分類タイプとも呼ばれます。

### Excel を使用したユーザー定義のスキーマの作成

次の例では、Parent/Category/Description スキーマを作成します。

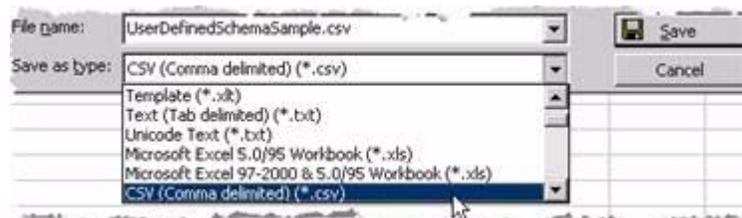
1. Excel を開いて、新規のワークシートを作成します。
2. 1 行目に、次の 3 つの列で構成されるヘッダーを作成します。

	A	B	C
1	Parent	Category	Description

3. スキーマ階層を各列に適切に入力します。**Parent** 列では、ツリーの最上位ノードは空白のままにできます。一方、子ノードはすべて入力する必要があります。

	A	B	C
1	Parent	Category	Description
2		UD_Cat1	User Defined Schema Category One
3	UD_Cat1	UD_Cat1_Child1	User Defined Schema Category One Level Two Child One
4	UD_Cat1_Child1	UD_Cat1_Child1_Commodity	User Defined Schema Category One Level Two Child One Commodity
5	UD_Cat1	UD_Cat1_Child2	User Defined Schema Category One Level two Child Two
6	UD_Cat1_Child2	UD_Cat1_Child2_Commodity	User Defined Schema Category One Level Two Child Two Commodity
7		UD_Cat2	User Defined Schema Category Two
8	UD_Cat2	UD_Cat2_Child1	User Defined Schema Category Two Level Two Child One
9	UD_Cat2_Child1	UD_Cat2_Child1_Commodity	User Defined Schema Category Two Level Two Child One Commodity
10	UD_Cat2	UD_Cat2_Child2	User Defined Schema Category Two Level two Child Two

4. ワークシートをカンマ区切りファイルとして保存するために、「ファイル」メニューから「名前を付けて保存...」を選択します。



5. 「ファイルの種類」リストから「CSV (カンマ区切り) (\*.csv)」を選択します。
6. ファイル名を入力し、「保存」をクリックします。  
作成したスキーマ・ファイルがカンマ区切りファイルとして保存されます。

## テキスト・エディタを使用したユーザー定義のスキーマの作成

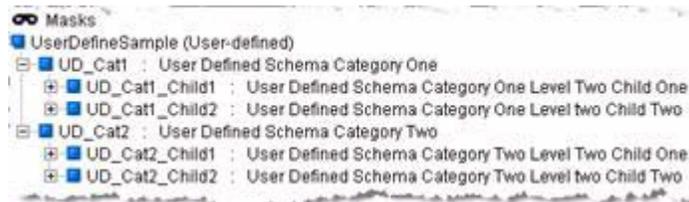
テキスト・エディタを使用して、前述の項の説明と同じ情報が格納されたカンマ区切りファイルを作成できます。このファイルの1行目には、次のヘッダー情報を必ず入力する必要があります。

Parent, Category, Description

分類にコードがないか、カテゴリがコードである場合は、より単純な Parent/Category スキーマを同じ方法で作成できます。

## ユーザー定義の分類タイプの作成

マスター分類ファイルとして作成したカンマ区切りファイルを選択しますが、他のタイプの場合と同様に、新しいユーザー定義の分類タイプを作成します。次の例に、ユーザー定義のスキーマの表示を示します。

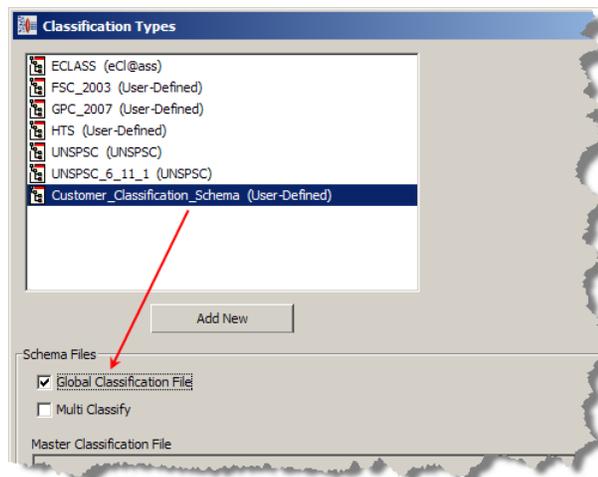


分類タイプの作成の詳細は、5-3 ページの「[分類タイプの作成](#)」を参照してください。

## グローバル分類スキーマ

ユーザー定義の分類スキーマを作成する場合は、このスキーマをグローバルにできます。これにより、分類スキーマの単一ユーザーが更新でき、変更内容は、そのユーザー定義の分類スキーマを使用しているすべてのデータ・レンズで使用可能になります。

新しいユーザー定義の分類タイプを追加する場合は、「**Global Classification File**」チェック・ボックスを選択してください。



分類タイプの作成の詳細は、5-3 ページの「[分類タイプの作成](#)」を参照してください。

## 分類システムのサポート

Knowledge Studio は、UNSPSC 部品分類システムのベンダー固有のレベル（レベル 5）に対する機能拡張をサポートしています。これらの機能拡張の作成方法の詳細は、コンサルタントにお問い合わせください。

## データ・レンズ分類システムのファイル形式

企業に独自または内部的に開発した分類システムがある場合、Knowledge Studio には、これらのスキーマをインポートできる形式があります。この形式では、最高 5 レベルの分類階層を使用できます。これらの機能拡張の作成方法の詳細は、コンサルタントにお問い合わせください。



---

## スマート用語集

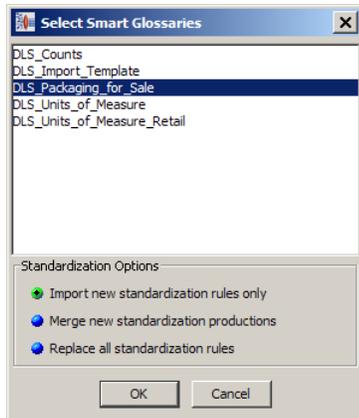
スマート用語集は、他のデータ・レンズにインポートできるセマンティック・ナレッジ（フレーズおよび用語）のセットです。

## スマート用語集のインポート

スマート用語集のインポートは優れた再利用機能です。スマート用語集をインポートすると、スマート用語集に含まれているフレーズおよび用語のルールがデータ・レンズですばやく効率的に取得されます。スマート用語集が修正され、その用語集を再度インポートしても、データ・レンズの用語およびフレーズのルールは（インポート以降に修正したかどうかに関係なく）変更されません。この項で後述するように、選択したオプションによっては、一部の標準化ルールに影響する可能性があります。

次の手順を使用して、新規または既存のデータ・レンズにスマート用語集をインポートできません。

1. 「File」メニューから「Import Smart Glossaries」をクリックします。



2. リストされているスマート用語集を1つ以上選択します。[Ctrl] キーを使用すると、リストで連続していない項目を複数選択できます。
3. オプションを使用して、次のいずれかの標準化オプションを選択します。

- **Import new standardization rules only**

このオプションは、ターゲット・データ・レンズにすでに存在するルールに、スマート用語集から新規の用語およびフレーズのルールや新規の標準化ルールをマージします。そのスマート用語集を再度インポートしても、ターゲット・データ・レンズに適用した変更は上書きされません。これがデフォルトです。

- **Merge new standardization productions**

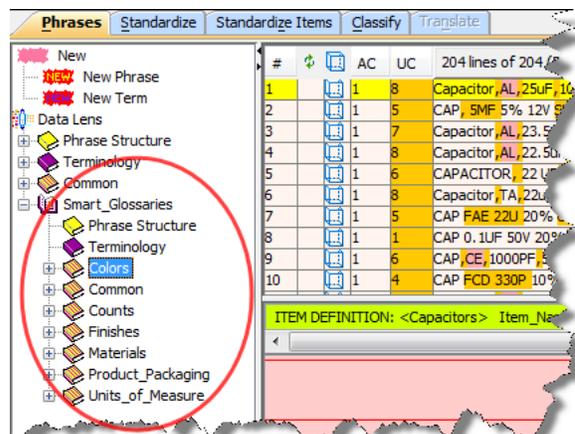
このオプションは、ターゲット・データ・レンズへの前回のインポート以降にスマート用語集に追加された新規標準化ルールをインポートし、新規標準化生成結果をターゲット・データ・レンズに追加します。そのスマート用語集を再度インポートしても、ターゲット・レンズの標準化生成結果への変更内容は保持されます。

- **Replace all standardization rules**

このオプションは、ターゲット・データ・レンズに対する標準化ルールのグローバル変更を実装します。

4. 「OK」をクリックします。

スマート用語集がデータ・レンズにインポートされ、ナレッジが適用されます。



## スマート用語集の作成

次の手順を使用して、新しいスマート用語集を作成できます。

---

**注意：** インポート可能なスマート用語集の作成は高度な機能です。新しいスマート用語集のインポート時にデバッグできないエラー・メッセージが表示された場合は、コンサルタントにお問い合わせください。

---

1. スマート用語集として指定するデータ・レンズをオープンします。
2. 「Data Lens」メニューで、「Data Lens Options」をクリックします。  
「Data Lens Options」ダイアログが表示されます。
3. 「Importable Lens」チェック・ボックスを選択して、このスマート用語集を他のデータ・レンズにインポートできるようにします。

---

**注意：** インポート可能なレンズ機能をアクティブ化するには、コンサルタントに問い合わせる必要があります。

---

4. 「OK」をクリックします。
5. データ・レンズを Oracle DataLens Server にチェックインします。  
データ・レンズが他のデータ・レンズにスマート用語集としてインポート可能になります。

## 付属のスマート用語集

この項では、ソフトウェア・リリースに付属の Enterprise DQ for Product スマート用語集について説明します。スマート用語集のファイルは、DLS\_接頭辞で識別されます。スマート用語集では品目定義は使用されていませんが、品目定義を使用するデータ・レンズには品目定義がインポートされます。

すべてのスマート用語集は、多数のデータ・セットで最も一般的で関連性のある形式を認識できるように、多種多様なデータについて広範なテストが実施されています。ただし、特定のデータについては、目的に対して結果が正しいことを保証するために、SME が認識の出力をレビューする必要があります。

## 色

色のスマート用語集 (DLS\_Colors) は、色と色系統をすばやく認識できるように設計されています。色は色系統に編成されます。基本色の標準化タイプが用意されているため、色の用語を色系統に標準化できます。たとえば、セルリアンという色は、青色系統に標準化されます。

基本色の標準化タイプは、各色を 11 の色系統のいずれかに標準化します。

## 数

数のスマート用語集 (DLS\_Counts) は、特定の品目 (脚、コンセントなど) の数をすばやく認識できるように設計されています。

数のスマート用語集は、ドメイン (電子部品、小売、照明、家電製品など) に表示される様々なタイプの、数をカウントする品目を認識します。このスマート用語集は、小さい数 (2 など) から大きい数 (12,000 など) の整数と、「one」から「twelve」までの整数のアルファベット表記を認識します。

次に、認識される形式の例を示します。

- 3-way
- 5 door
- three tier
- 2sided

このスマート用語集には、「pair」、「item」、「count」などの DLS\_Product\_Packaging にあるような用語はありません。

このレンズで使用される用語のバリエーションは、妥当な略語と、おそらくはスペルミスです。

既知のあいまい表現が 1 つ特定されています。データに含まれる部品番号の後に、DLS\_Counts の数をカウントする用語が続く場合 (例: UPC: 123123123 door) は、部品番号が正しく認識されません。これは、DLS\_Counts から不適切なフレーズ構造ルールを削除することで簡単に修正できます。

## 販売用パッケージ設定

販売用パッケージ設定のスマート用語集 (DLS\_Product\_Packaging) は、商品の販売用パッケージ設定の説明に使用される一連の共通パッケージ、数量および単位を認識します。データ・レンズは、事務用品、ティッシュ、バイオ廃棄物処理製品、おもちゃ、紙製品、家庭用品、食品、園芸用品、工具など、パッケージ商品の幅広い市場から選択した製品に対してテストされています。

販売用パッケージ設定のスマート用語集は、様々なタイプのパッケージ設定とそれらのパッケージ設定タイプのすべての組合せ (1 箱当たりの缶数、1 カートン当たりの箱数、1 カートン当たりの缶数、1 ケース当たりの箱数など) を認識します。3 桁区切り付き / なしの数字表記の数量 (12,000 や 12000)、「one」から「twelve」までのアルファベット表記の数字、「pair」、「dozen」、「ream」、「gross」などのアルファベット表記の数量を認識します。

このスマート用語集は、2つのレベルのパッケージ設定を認識します。

- パッケージ当たりの単位 (18 缶で 1 箱など)
- コンテナ当たりのパッケージ (28 箱で 1 ケースなど)

すべての単位は、数字に標準化されます。

データにテキストベースの数量用語が必要な場合、またはパッケージ・タイプがスマート用語集ルール・セットに含まれていない場合は、ニーズに対応するように既存のルールを簡単に修正できます。

このスマート用語集は、パッケージ設定の単位を最大限認識するように設計されていますが、品目の中には、あるドメインではパッケージを表し、別のドメインでは品目や製品を表すものがあります。たとえば、紙製品は、パッド状にパッケージされた複数の用紙 (メモ帳) として製造されることがあります。パッドがパッケージではなく品目を表す場合は、パッケージ・タイプとしてのパッドを除外するように、ターゲット・データ・レンズを簡単に修正できます。

このスマート用語集では、価格設定情報は認識しません。価格は、一般的に入力データから除外されます。販売用パッケージ設定情報が含まれているデータに価格情報を格納する場合は、これらの 2 つのタイプの情報を区別するために、価格設定情報とパッケージ設定数量情報の両方が含まれているフレーズ・ルールを追加します。ベスト・プラクティスは、データ・セットから価格情報を排除することです。

このスマート用語集では、重量で定量化される単位 (1 箱当たり 14 オンスなど) は認識しません。一般的に数式スタイルの摘要 (例 : bags per box [=] 15 boxes per case [=] 12) は認識しません。

## 単位

単位のスマート用語集 (DLS\_Units\_of\_Measure) は、ユーザーが速やかに開始して最小の労力で最も一般的な単位を検出できるようにします。

スマート用語集では、多数のターゲット市場に対応するように、広範囲にわたる一般的な単位を認識します。

- 時間
- 長さや距離
- 電圧
- 抵抗
- 許容範囲
- データとデータ転送速度
- 音
- ワイヤ・ゲージ
- 温度

単位のスマート用語集は、次のような同じタイプの単位間で換算が必要な場合は単位換算にも対応します。

- 長さや距離: メートルからフィートまたはインチ
- 容積: リットルからガロンまたはクォート
- 電力: キロワットからワット
- 抵抗: オームからキロオーム

単位のスマート用語集には、用語またはその略語について回避できない多くの競合が存在します。これは、この項で説明するように、データに必要な単位を一意に識別するために、インポート後にいくつかのルールを削除するか、(追加のルールを使用するか品目定義を使用して) ルール・セットを追加する必要があることを意味します。

同じ略語を他の用語と共有するあいまい表現を回避するために、標準的な略語であってもスマート用語集に登録されていないものが数多くあります。

- M: M バイトまたは M ビットの略語として使用されますが、メートルにのみ適用されます。
- W: 幅やワットの略語として使用されますが、登録されていません。
- L: 長さやリットルの略語として使用されますが、登録されていません。
- F: 華氏とファラッドの両方の略語として使用されますが、華氏の略語としてのみ登録されています。

さらに、データには単位として検出されることがある製品番号や製品コードが含まれている可能性があります。これは、ルールからの未使用生成結果の削除、行の最初と最後の引用符の削除、コンテキスト内で品目を区別するための品目定義の使用などの手法を使用して、ターゲット・データ・レンズの最小のリファクタリングで修正できます。

また、このスマート用語集では C と F が温度スケールの摂氏および華氏の略語として認識されますが、これは予想外の結果を引き起こす可能性があります。これは、必要でない略語を削除したり、品目定義で値ロジックを使用して無効な温度範囲を除外するなどの手段で簡単に修正できます。値ロジックの設定については、コンサルタントにお問い合わせください。

## 単位（小売）

単位（小売）のスマート用語集（DLS\_Units\_of\_Measure\_Retail）には、次の項で詳細に説明されているように、一般的に小売データに存在する単位のみが含まれています。ファラッド、ピコファラッド、ジュール、マイクロヘンリ、awg 値などのより専門的な単位を認識するためには、標準的な単位のスマート用語集（DLS\_Units\_of\_Measure）をインポートする必要があります。あまり一般的でない単位について特別な用語やフレーズのルールを追加せずに小売データの単位を認識するためには、このスマート用語集を使用します。

このスマート用語集が認識する単位のタイプの中には、次のものがあります。

- アンペア数
- データ転送速度
- エネルギー
- 長さ
- 電力
- 温度
- 時間
- 電圧
- 容積
- 重量

このスマート用語集は、DLS\_Units\_of\_Measure と併用しないように設計されています。DLS\_Units\_of\_Measure を以前にインポートしたデータ・レンズに DLS\_Units\_of\_Measure\_Retail をインポートすると、DLS\_Retail\_Units\_of\_Measure の階層構造が DLS\_Units\_of\_Measure の階層構造と一体化する可能性があります。

また、このスマート用語集では C と F が温度スケールの摂氏と華氏の略語として認識されますが、これは予想外の結果を引き起こす可能性があります。これは、必要でない略語を削除したり、品目定義で値ロジックを使用して無効な温度範囲を除外するなどの手段で簡単に修正できます。値ロジックの設定については、コンサルタントにお問い合わせください。



---

## クライアント・ソフトウェアのインストール

---

Enterprise DQ for Product は、Java Web Start を使用して、クライアント・デスクトップにソフトウェアの現行バージョンを初期インストールして保守します。このプロセスでは、Oracle DataLens Server にアクセスし、接続を開始してソフトウェアをダウンロードする必要があります。

Java Web Start を使用して Enterprise DQ for Product クライアント・アプリケーションをダウンロードおよびインストールするには、Oracle DataLens Server のインストール・ページを参照し、次の手順に従います。

1. Java SE Runtime Environment (JRE) 6 Update 21 がインストールされていることを確認します。JRE のダウンロードとインストール方法については、次の URL にアクセスしてください。

<http://www.oracle.com/technetwork/java/javase/downloads/index.html>

2. Microsoft Internet Explorer を起動します。

3. 接続を開始し、次の URL にアクセスしてクライアント・ソフトウェアをダウンロードします。

`http://server:2229/datalens/datalens.jnlp`

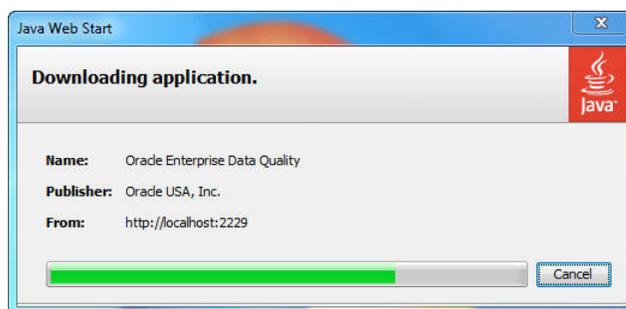
`server` は、Oracle DataLens Server のホスト名です。

---

**注意：** アプリケーション・サーバーに対して 2229 以外のポート番号を設定している場合は、そのポート番号を次の URL 内で使用して Oracle DataLens Server にアクセスし、クライアント・アプリケーションをダウンロードする必要があります。

---

アプリケーションのダウンロードと検証が開始されます。



---

**注意：** .jnlp ファイルがプログラムに関連付けられていないことを示す  
ファイル・ダウンロード・メッセージが表示された場合は、サポートされて  
いる JRE がインストールされていません。「Cancel」をクリックしてステッ  
プ 1 に戻ります。

---

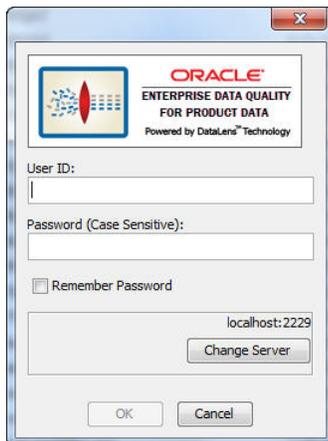
検証が完了すると、インストールが開始されます。Enterprise DQ for Product ファイルは信  
頼できるソースによってデジタル署名されているため、次のセキュリティ警告が表示され  
ます。



**ヒント：** このセキュリティ・ダイアログが今後表示されないようにするに  
は、「Always trust content from this publisher」チェック・ボックスを選択  
します。

4. 「Run」をクリックしてインストールを完了します。

Oracle Enterprise Data Quality for Product Data のログイン・ダイアログが表示されます。



# B

---

## 正規表現

正規表現では、文字パターン照合を使用して必要な情報を検索し、取得します。正規表現は、**Knowledge Studio** で用語ルールを作成する際に最も頻繁に使用します。

正規表現を使用するには、その構文を習得する必要があります。正規表現では、特殊文字であるワイルドカードを使用して、他の文字列の範囲を照合します。用語ルール内の正規表現は、フォワード・スラッシュで囲みます。

## 正規表現で使用する特殊文字

次の表に、正規表現で使用する特殊文字と式の例をいくつか示します。

ワイルドカード またはメタ文字	説明と例
.	ドット文字はすべての単一文字と一致します。 たとえば、用語ルールの正規表現 <code>/a.b/</code> は、 <code>a5b</code> のように、 <code>a</code> の後に任意の単一文字が続き、その後に <code>b</code> が続くすべてのテキストと一致します。
*	アスタリスクは、直前のパターンや文字と 0 (ゼロ) 回以上一致します。 たとえば、 <code>/fo*/</code> は、次のテキスト・フラグメントと一致します。 <code>f</code> 、 <code>fo</code> 、 <code>foo</code> 、 <code>fooo</code> ピリオドとアスタリスクを組み合わせた <code>/a.*b/</code> は、 <code>a5b</code> 、 <code>a55b</code> 、 <code>a123b</code> などと一致します。
+	プラス記号は、直前のパターンや文字と 1 回以上一致します。 たとえば、 <code>/ca+r/</code> は <code>car</code> 、 <code>caar</code> および <code>caaar</code> と一致しますが、 <code>cr</code> とは一致しません。
?	疑問符文字は、直前のパターンや文字と 0 (ゼロ) 回または 1 回一致します。 たとえば、 <code>/ca?r/</code> は <code>car</code> および <code>cr</code> の両方と一致しますが、 <code>caar</code> とは一致しません。
{n}	中カッコは、直前の文字やパターンの <code>n</code> 個のインスタンスと正確に一致させる場合に使用します。 たとえば、 <code>/x{2}/</code> は <code>xx</code> と一致します。 <b>注意:</b> アプリケーションでは、空白で区切られたテキストや文字列と、識別可能な空白がない、他の文字列に埋め込まれたテキストや文字列を区別するために中カッコを使用します。
{n,m}	この形式の中カッコは、直前の文字やパターンと <code>n</code> 回～ <code>m</code> 回 ( <code>n</code> は <code>m</code> より小さい) 一致させる場合に使用します。 <code>m</code> がいない場合、このパターンは <code>n</code> 回以上照合されます。 たとえば、 <code>/x{2,3}/</code> は <code>xx</code> および <code>xxx</code> と一致します。
[...]	大カッコを使用すると、カッコ内の文字のいずれかと一致します。アルファベット文字の範囲は、ハイフンを使用して照合できます。 たとえば、 <code>/[xyz]/</code> は、 <code>x</code> 、 <code>y</code> または <code>z</code> のいずれかと一致します。また、 <code>/[xyz]/+</code> は、 <code>x</code> 、 <code>xx</code> 、 <code>y</code> 、 <code>yy</code> などと一致します。 大カッコ内の文字の範囲は、ダッシュ (-) を使用して定義できます。たとえば、 <code>[a-z]</code> は、いずれかの小文字と一致し、 <code>[A-Z]</code> は、いずれかの大文字と一致します。ダッシュを使用して文字の範囲を定義する場合は、アルファベット順または数字順で、最初の文字を 2 番目の文字より前に指定する必要があります。 たとえば、 <code>[0-9]</code> は有効ですが、 <code>[9-0]</code> は無効です。
(...)	丸カッコは、文字列をグループ化する場合に使用します。 たとえば、 <code>(cars?) bus</code> は、 <code>car</code> 、 <code>cars</code> または <code>bus</code> と一致します。 <b>注意:</b> この丸カッコは <code>(?...)</code> と同じです。
x y	パイプ ( ) 文字は、 <code>x</code> または <code>y</code> のいずれかと一致します。 <code>x</code> または <code>y</code> は文字のブロックです。 たとえば、 <code>car bus</code> は、 <code>car</code> または <code>bus</code> と一致します。

ワイルドカード またはメタ文字	説明と例
¥	<p>円記号には2つの意味があります。</p> <p>星 (*) やドット (.) など、通常は特別な意味を持つ文字を照合します (前述の説明を参照)。この場合、¥* は星文字と一致します。同様に、"¥." はドット文字と一致します。</p> <p>メタ文字を定義するために使用されます。通常、文字 w は、w と一致します。¥w は、空白で区切られていない一連の英数文字と一致します (次の説明を参照)。</p>
¥w	すべての英数文字またはアンダースコアと一致します。これは、[A-Za-z0-9_]と同じです。
¥W	英数文字またはアンダースコア以外のすべての文字と一致します。
¥d	<p>すべての数字と一致します。[0-9] と同じです。</p> <p>たとえば、/¥d+/ は、1つ以上の数字と一致します。</p> <p>たとえば、正の整数です。</p>
¥D	数字以外のすべて (空白も含む) と一致します。
¥s	タブやスペースなどの、すべての空白文字と一致します。
¥S	空白文字以外のすべての文字と一致します。
(?i)	<p>(?i) メタ文字は、照合の実行時に、直後のパターンで文字の大文字と小文字を区別しないことを示します。</p> <p>たとえば、(?i)car は、Car、car、cAR などと一致します。また、(?i)cars? は、Car、Cars、CarS などと一致します。</p> <p><b>注意:</b> この照合ルールと次の1つのルールの構文上の相違は、パターンが丸カッコ内にあることです。</p>
(?!pattern1)pattern2	<p>(?!...)... メタ文字は、最初のパターン pattern1 が存在しない場合に、2番目のパターン pattern2 を受け入れます。</p> <p>たとえば、/(?!x)car/ は car と一致しますが、xcar とは一致しません。</p> <p><b>注意:</b> pattern1 と pattern2 の両方が必須です。</p>

## 用語ルールで役立つ正規表現

### 年

```
[year] case-insensitive|
/[12][0-9][0-9][0-9]/
```

### 郵便番号

```
[zip] case-insensitive
/\d\d\d\d\d/ # 5 digit zip code
/\d\d\d\d\d-\d\d\d\d\d/ # 9 digit zip code
```

### 名

```
[first_name] case-insensitive|
{[A-Za-z]+} # a name surrounded by white space
```

### 番地

```
[street_name] case-insensitive
{!(?!dr|ave|ln|ct|st)[A-Za-z]+} # street name not starting with dr, ave, ln, ...
{\d[Rr][Dd]} # 3rd, ...
{\d\d[Rr][Dd]} # 23rd, ...
{\d\d\d[Rr][Dd]} # 103rd, ...
{\d[tT][hH]} # 5th, ...
{\d\d[tT][hH]} # 25th, ...
{\d\d\d[tT][hH]} # 105th, ...
{\d[sS][tT]} # 1st, ...
```

正規表現の詳細は、『Perl for Dummies』（Paul Hoffman 著）または『Mastering Regular Expressions』（Jeffrey Friedl 著）を参照してください。