

IBM SPSS Modeler Entity  
Analytics 15 ユーザー ガイド



注：この情報とサポートされている製品をご使用になる前に、「注意事項」（p.72）の一般情報をお読みください。

本版は ，および新版で指示されるまで後続するすべてのリリースおよび変更に対して適用されます。

Adobe 製品のスクリーンショットは Adobe Systems Incorporated の許可を得て転載しています。

Microsoft 製品のスクリーンショットは Microsoft 社の許可を得て転載しています。

Licensed Materials - Property of IBM

U.S. Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

---

# はじめに

IBM® SPSS® Modeler は、IBM Corp. が開発した企業強化用のデータ マイニング ワークベンチです。SPSS Modeler を使用すると、企業はデータを詳しく調べることで顧客および一般市民とのリレーションシップを強化することができます。企業は、SPSS Modeler を使って得られた情報に基づいて利益をもたらす顧客を獲得し、抱き合わせ販売の機会を見つけ、新規顧客を引き付け、不正を発見し、リスクを減少させ、政府機関へのサービスの提供を改善することができます。

SPSS Modeler の視覚的インターフェイスを使用すると、特定ビジネスの専門知識を適用し、より強力な予測モデルを実現し、解決までの時間を短縮します。SPSS Modeler では、予測、分類、セグメント化、および関連性検出アルゴリズムなど、さまざまなモデル作成手法を提供しています。モデルを作成した後は、IBM® SPSS® Modeler Solution Publisher により、企業全体の意思決定者やデータベースにモデルを配布することが可能になります。

## IBM Business Analytics について

IBM Business Analytics ソフトウェアは、意思決定者がビジネス パフォーマンスを向上させるために信頼する完全で、一貫した正確な情報を提供します。ビジネス インテリジェンス、予測分析、財務実績および戦略管理、および分析アプリケーションの包括的なポートフォリオを利用することによって、現在の実績を明確、迅速に理解し、将来の結果を予測することができます。豊富な業界のソリューション、実績ある実例、専門サービスと組み合わせ、さまざまな規模の組織が、高い生産性を実現、意思決定を自信を持って自動化し、より良い決定をもたらします。

このポートフォリオの一部として、IBM SPSS Predictive Analytics ソフトウェアを使用する組織は、将来のイベントを予測し、その洞察に基づいて積極的に行動し、より優れた業績を実現することができます。全世界の企業、政府、学術分野のお客様が IBM SPSS の技術を活用し、不正行為を減少させ、リスクを軽減させながら、顧客の獲得、保持、成長において、競争優位を高めることができます。IBM SPSS ソフトウェアを日々の業務に取り入れることによって、組織は業務目標を達成し、大きな競争的優位を獲得することができるよう、意思決定を方向付け、自動化することができるようになります。お問い合わせは、<http://www.ibm.com/spss> を参照してください。

## テクニカル サポート

お客様はテクニカル サポートをご利用いただけます。IBM Corp. 製品の使用方法、または対応するハードウェア環境へのインストールについてサポートが必要な場合は、テクニカル サポートにご連絡ください。テクニカ

ル サポートの詳細は、IBM Corp. Web ページ <http://www.ibm.com/support> を参照してください。ご本人、組織、サポートの同意を確認できるものをご用意ください。

---

# 内容

<b>1 エンティティ分析</b>	<b>1</b>
エンティティ分析について	1
エンティティ分析と予測分析	3
<b>2 IBM SPSS Modeler によるエンティティ分析</b>	<b>4</b>
IBM SPSS Modeler によるエンティティ分析の活用	4
段階 1: ソース データを SPSS Modeler に読み込む	5
段階 2: リポジトリを作成する	6
段階 3: SPSS Modeler をリポジトリに接続する	7
段階 4: リポジトリのフィーチャに入力フィールドをマッピングする	7
段階 5: データをリポジトリにエクスポートして競合を解決する	9
段階 6: 解決した識別の分析	11
段階 7: リポジトリに対して新しいケースを解決する	12
段階 8: アラートを生成する	14
<b>3 エンティティ分析のタスク</b>	<b>15</b>
タスクについて	15
エンティティ リポジトリの設定 (EA エクスポート ノード)	15
エンティティ リポジトリ	15
データ ソースへの接続	16
リポジトリの作成	17
入力フィールドのフィーチャへのマッピング (EA エクスポート ノード)	20
フィールド マッピングの表示 (EA エクスポート ノード)	24
エンティティ リポジトリの構成	25
データ ソース マッピングの表示	26
リポジトリのフィーチャの更新	26
フィーチャの追加または編集	29
エンティティ タイプの管理	31
エンティティの一致に対するしきい値の設定	34
リポジトリ設定の再利用	35
設定の変更の保存	35
設定ウィンドウを閉じる	35

解決した識別子を分析するには (Entity Analytics (EA) 入力ノード) . . . . .	36
データソースの選択 . . . . .	36
データフィールドの名前の変更 . . . . .	37
データフィールドのデータ型情報の設定 . . . . .	37
ノードのストリームへの追加 . . . . .	38
リポジトリに対して新しいケースを解決する (ストリーミング EA ノード) . . . . .	38
入力フィールドのフィーチャへのマッピング (ストリーミング EA ノード) . . . . .	40
フィールド マッピングの表示 (ストリーミング EA ノード) . . . . .	41
データソースの表示 (ストリーミング EA ノード) . . . . .	42
ストリーミング EA ノードの出力 . . . . .	44
他の IBM SPSS 製品との IBM SPSS Modeler Entity Analytics の使用 . . . . .	44
管理タスク . . . . .	44
ポート割り当ての設定 . . . . .	45
リポジトリ データベースの管理者の資格情報の管理 . . . . .	46
リポジトリ ストレージ ディレクトリの移動 . . . . .	47
日付/時刻およびタイムスタンプのフィールドのストリーム オプションの設定 . . . . .	47
同じ Windows システムでの SPSS Modeler クライアントと SPSS Modeler Server との IBM SPSS Modeler Entity Analytics の実行 . . . . .	48
エンティティ リポジトリの消去 . . . . .	48
エンティティ リポジトリの削除 . . . . .	49
リポジトリに接続できない場合にリポジトリを削除する . . . . .	50

## 4 実際のエンティティ分析 51

この例について . . . . .	51
元のモデル . . . . .	51
エンティティ分析の追加 . . . . .	56
ソース データをリポジトリに取得する . . . . .	56
解決した識別の読み込み . . . . .	58
エンティティ分析出力と元のモデルとの比較 . . . . .	67
要約 . . . . .	71

## 付録

A 注意事項 72

索引 75





# エンティティ分析

## エンティティ分析について

IBM® SPSS® Modeler Entity Analytics が新しい次元を IBM® SPSS® Modeler の予測分析に追加します。予測分析は過去のデータから将来の行動を予測しようとするのに対し、エンティティ分析ではレコードの中でアイデンティティの競合を解決することで現在のデータの干渉性と一貫性を改善することに焦点を当てます。アイデンティティは、個人、組織、オブジェクトまたは曖昧さの存在する他のエンティティとなります。アイデンティティの解決は、顧客関係の管理、不正行為の検出、マネーロンダリング防止、国内および国際的なセキュリティなどのさまざまなフィールドにおいて重要になります。

2 つの異なるソースに以下の顧客レコードがあり、彼らが同一人物なのかまたは異なる人々を示しているかどうか分からないと仮定します。

### ソース 1

Record no.: 70001  
Name: Jon Smith  
Address: 123 Main Street  
Tax Reference: 555-00-1111  
Driv.License: 0001133107  
Cred.Card: 10229127

### ソース 2

Record no.: 9103  
Name: JOHNATHAN Smith  
Date of Birth: 06/17/1934  
Telephone: 555-1212  
Cred.Card: 10229128  
Email: jls@mail.com  
IP address: 9.50.18.77

2 つのレコード間のデータの完全一致は存在しません。ただし、3 番目のソースを導入すると、いくつか共通の属性が見つかります。

### ソース 3

Record no.: 6251  
Name: Jon Smith

Telephone: 555-1212  
Driv.License: 0001133107  
Cred.Card: 10229132

運転免許番号はソース 1 とソース 3 でリンクしますが、電話番号はソース 2 とソース 3 でリンクしています。そのため、3 つのソースすべてが同じ人物を示していることを合理的に確認することができます。

容易に区別できない場合はどうすればよいでしょう?意思決定の基準となるデータが非常に少ない場合があります。以下の 2 つのレコードについて考えてみましょう。

#### ソース 4

Record no.: S45286  
Name: John T Smith Jr  
Address: 456 Main Street  
Telephone: 703-555-2000  
Date of birth: 03/12/1984

Record no.: S45287  
Name: John T Smith  
Address: 456 Main Street  
Telephone: 703-555-2000  
Driv.License: 009900991

明らかにこれは前のレコードと同じ Smith 氏ではありません。この違いにより、このレコードを除外することができます。ただしまだ問題があります。同じデータソースの二つの異なるレコードは、同じ人物に関連するよう考えられます。これらは重複したレコードなのでしょう吗?おそらく別のソースからより詳細な情報をもたらす関連レコードがないかぎり、確かではありません。

#### ソース 5

Record no.: 769582-2  
Name: John T Smith Sr  
Address: 456 Main Street  
Telephone: 703-555-2000  
Driv.License: 009900991  
Date of birth: 06/25/1959

これにより、問題が解決します。ソース 4 の 2 つのレコードは重複していませんが、実際は同じ名前、同じ住所、同じ電話番号を使用している父親と息子です。手動によるシステムでは、アイデンティティを解決した 2

つのレコードを見つけるために数週間かかる場合があります。自動化されたエンティティ分析システムで、解決時間が劇的に短縮します。

## エンティティ分析と予測分析

すべてのデータが完全かつ明白なレコードの単一のソースで構成されている場合、IBM® SPSS® Modeler は比較的容易に任意のアイデンティティの競合を解決することができます。予測分析のみを利用した場合、SPSS Modeler にデータを読み込み、処理を実行して信頼できる結果を取得することが可能でした。

ただし実際、実態は非常に異なります。データは不完全で、曖昧なことが多く、さまざまなデータ ソースに分散しており、フィールドがほとんど特服していないさまざまな属性を記録しています。エンティティ分析の値の一部は、さまざまなデータ ソースから単一の集約ストレージ データ、リポジトリに収集にあります。エンティティ分析システムが、同じ人物または組織に由来するレコードに一意の識別子を追加し、データを詳細に検証して競合を解決します。

次の表は、2 つの分析の間の相違点について示しています。

テーブル 1-1  
予測分析とエンティティ分析の相違

特徴	予測分析	エンティティ分析
学習	比較的小さいセットおよび数値の範囲に基づく	名前、住所、などの大きなセットに基づく
学習に使用するデータ	サブセット (学習区分)	使用されるすべてのデータ
一般化	アルゴリズムは簡潔なモデルを形成するために学習データ全体を一般化します。	データは、エンティティのマッチングと関係の検出に適した構造に永続化されます。
不正検出	レコードは、それらが不正アプリケーションの典型的な特徴を持っている場合に不正の可能性があるとフラグが付きま	既知の不正レコードに関連する場合、または同じ人物であるが異なる ID である場合に不正の可能性のあるものとしてフラグが付きま

# IBM SPSS Modeler によるエンティティ分析

## IBM SPSS Modeler によるエンティティ分析の活用

データに識別上の問題があることが分かった場合、IBM® SPSS® Modeler Entity Analytics を使用してどのように対処したらよいでしょうか?以下は提案されている手順ですが、特定の要件に合わせて手順を変える必要があります。

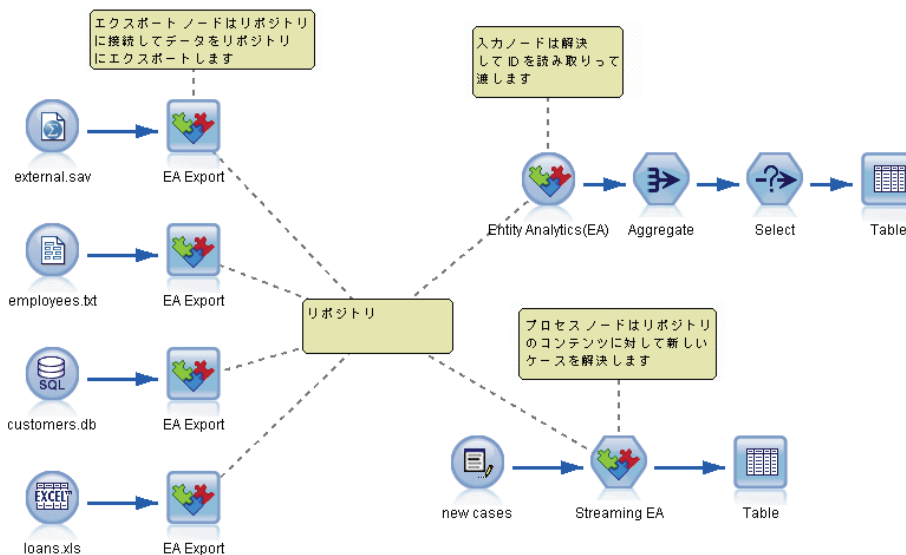
- ソース データを IBM® SPSS® Modeler に読み込む
- データを保存するリポジトリを作成する
- SPSS Modeler をリポジトリに接続する
- リポジトリのフィーチャにデータ フィールドをマッピングする
- データをリポジトリにエクスポートして識別を解決する
- 解決した識別を分析する
- リポジトリに対して新しいケースを解決する
- 必要なアラート（バッチまたはリアルタイム）を生成する

この時点で、SPSS Modeler がどのように機能するかについて一部知っておく必要があります。SPSS Modeler は、さまざまなノードを通過するデータのストリームのグラフィカル表示に基づいた、ユーザーフレンドリーなツールです。各ノードは、ワークフローの特定の段階を示します。

SPSS Modeler にはさまざまなノードがあり、標準的なデータ マイニング機能をすべてカバーしています。IBM SPSS Modeler Entity Analytics は特にエンティティ分析で使用するためのノードを追加します。これらは、EA エクスポート ノード、エンティティ分析 (EA) 入力ノード、およびストリーミング EA プロセス ノードです。

以下の図は、そのプロセスを示しています。

図 2-1  
エンティティ分析のプロセス



## 段階 1 : ソース データを SPSS Modeler に読み込む

まず、丸いアイコンによって SPSS Modeler で示されている入力ノードを使用して、データを SPSS Modeler に読み込みます。

図 2-2  
IBM SPSS Modeler 入力ノード



データは、テキストファイル、データベース テーブル、スプレッドシート、XML ファイルなど、SPSS Modeler でサポートされている形式であればどのようなものでも構いませんが、それぞれの形式で対応する SPSS Modeler 入力ノードが必要です。図では、データベース入力ノードを示しています。

各データ ソース ファイルには、各レコードを一意に識別する 1 つのフィールドが必要です。データ ソースにそのようなフィールドがない場合、SPSS Modeler のフィールドを簡単に追加できます。詳細は、3 章 p. 17 [一意のレコード識別子の追加](#) を参照してください。

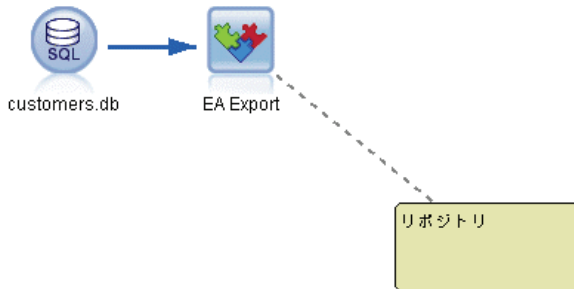
詳細は、3 章 p. 16 [データ ソースへの接続](#) を参照してください。

## 段階 2 :リポジトリを作成する

エンティティ分析の効果における焦点は、リポジトリ、つまりすべてのデータ レコードを収集する集約ストレージ領域です。

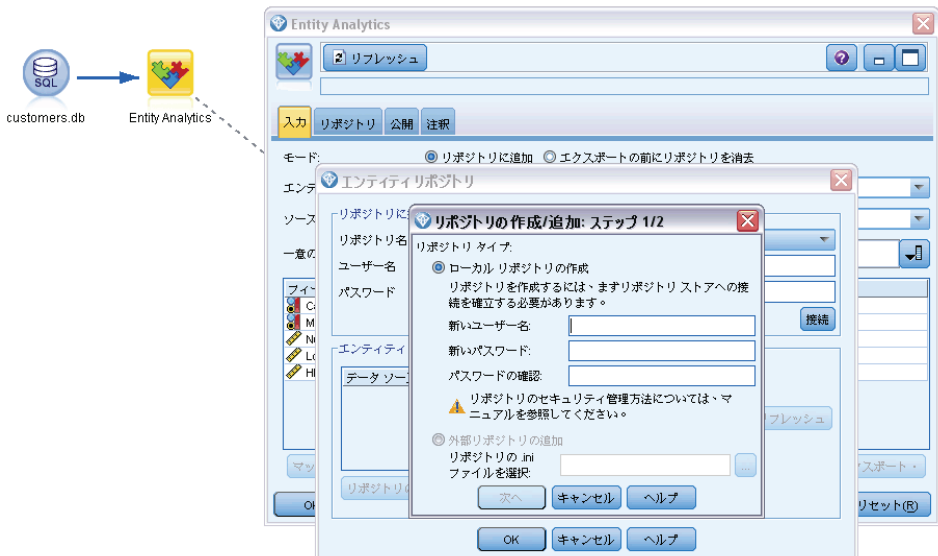
リポジトリを作成するには、まず正方形のアイコンで表される EA エクスポートノードにデータ ソースを接続します。

図 2-3  
リポジトリへの接続



エクスポート ノードから、新しいリポジトリ（または既存のリポジトリを選択）を作成し、エクスポートされたデータを受け取る準備をすることができます。

図 2-4  
リポジトリを作成する



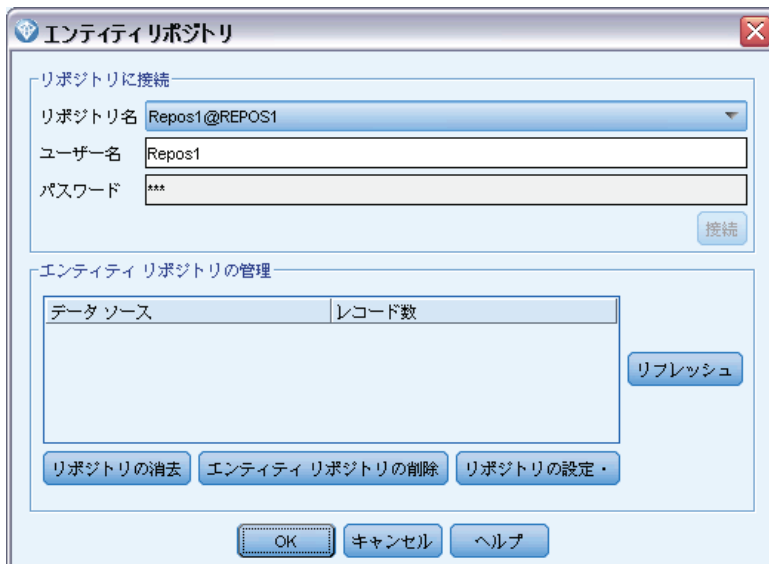
リポジトリ作成のプロセスについては、後で詳細に説明します。詳細は、3 章 p.15 エンティティ リポジトリの設定 (EA エクスポート ノード) を参照してください。

リポジトリを設定する場合、さまざまな方法でその内容を維持することができます。詳細は、3 章 p.25 エンティティ リポジトリの構成 を参照してください。

### 段階 3 :SPSS Modeler をリポジトリに接続する

リポジトリを作成したら、SPSS Modeler ストリームに接続します。

図 2-5  
リポジトリへの接続



詳細は、3 章 p.19 エンティティ リポジトリのオプション を参照してください。

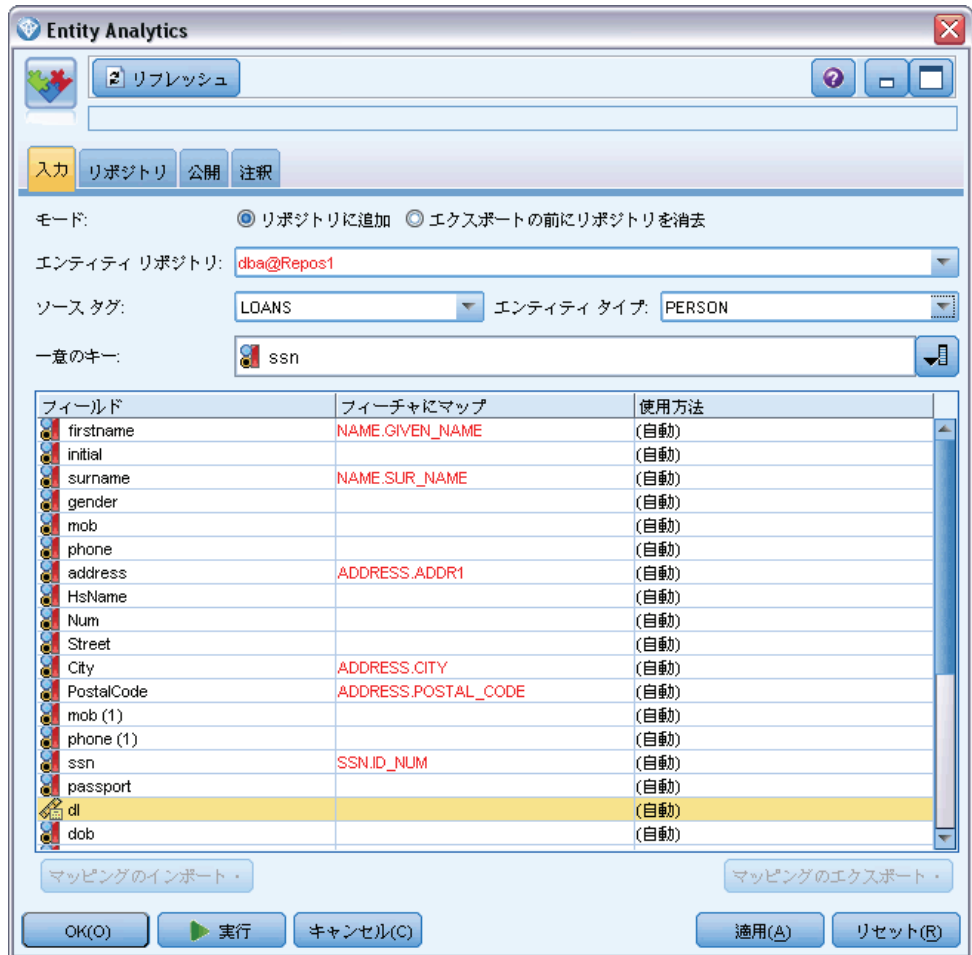
### 段階 4 :リポジトリのフィーチャに入力フィールドをマッピングする

データソースには、さまざまなエンティティ情報を含むことができます。多くのエンティティ データ ソースに共通する情報タイプと、特定のデータ ソース固有の情報タイプがあります。エンティティ リポジトリでは、これらのさまざまな情報の種類はフィーチャと呼ばれます。リポジトリは、標準としてさまざまなフィーチャを提供しており、独自のフィーチャを作成することもできます。

リポジトリ フィーチャは、エンティティ データ ソースで使用できる個別の情報タイプです。特定のデータソース固有のフィーチャ、（姓、名、生年月日など）さまざまなデータ ソースで使用されることができフィーチャがあります。フィーチャは通常、データ レコードのフィールド、またはデータベース テーブルの列に対応します。

リポジトリを作成して接続したら、入力データの 1 つのフィールドを一意キー フィールドとして指定し、後続の分析で使用します。また、リポジトリ内の対応するフィーチャに、入力データのフィールドをマッピングします。これは、異なるデータソースが同じ種類の情報を含むフィールドに異なる名前を使用する場合、リポジトリ内の重複を避けるためです。EA エクスポート ノードには、マッピングを作成することができるマッピング テーブルが用意されています。

図 2-6  
入力フィールドのフィーチャへのマッピング



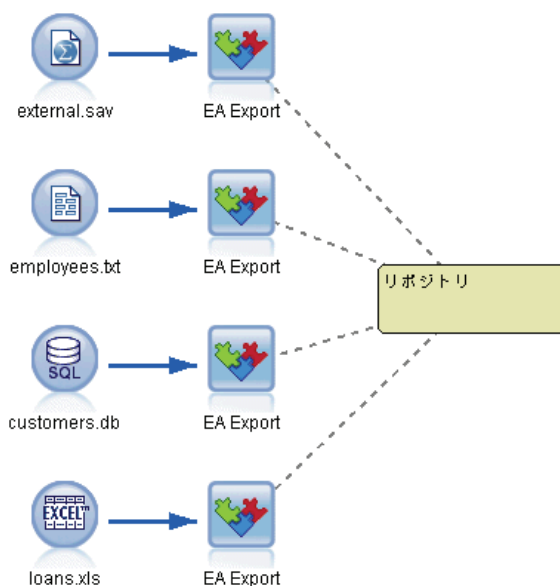


詳細は、3 章 p.20 入力フィールドのフィーチャへのマッピング (EA エクスポート ノード) を参照してください。

## 段階 5 :データをリポジトリにエクスポートして競合を解決する

各データ入力ノードでは独自の EA エクスポート ノードを必要とするため、データがさまざまなソースに分散される場合、このようなストリームとなります。

図 2-7  
データを複数のデータソースからリポジトリにエクスポート



データソースが複数ある場合、1 つ、いくつか、またはすべてのデータソースのレコードを読み取ることができます。エンティティ分析システムは、選択したレコードを分析し、それぞれに識別子フィールド `named $EA_ID` を追加します。以前曖昧な識別に関連する複数のレコードを解決できる場合、これらのレコードに追加される識別子は、リポジトリ全体で一意となります。システムはまた、レコードの作成元のデータソースを示すフィールドを追加します。

データ入力ノードを独自の EA エクスポート ノードに接続し、入力フィールドをリポジトリ フィーチャにマッピングし、ストリームを実行して SPSS Modeler のデータをリポジトリにエクスポートし、識別子の競合を解決する作業を一度の操作で行います。以上の作業がどのように行われるか示すために、4 つのデータソースに以下のレコードがあると想定します。

### 外部データ

名前	電話番号	信用リスク
Mike	555-1234	560
Joe	555-4567	780

### 従業員

名前	住所	電話番号
Michael	1234 5th Street	555-1234
Fred	543 1st Avenue	555-9876

### 顧客

名前	住所	貯金
Susan	1234 5th Street	\$1234
Joe	777 Oak Street	\$5

### ローン

名前	住所	電話番号	ローン
Sue	1234 5th Street	555-1234	\$10,000
Joseph	777 Oak Street	555-4567	\$50,000

ここで示すように、それぞれのデータソースを順番にリポジトリにエクスポートします。エクスポートすると、リポジトリが各レコードの解決を更新します。リポジトリでは、各レコードの前に識別子フィールド（\$EA-ID）とソース インジケータ フィールド（\$EA-SRC）が付き、レコードが生じたデータ ソースを示します。そのためこの例では、4 つのデータソースすべてをエクスポートすると、リポジトリの内容は次のようになります。

テーブル 2-1  
エクスポート後のリポジトリの内容の例

\$EA-ID	\$EA-SRC	名前	電話番号	住所	信用リスク	貯金	ローン
1	従業員	Michael	555-1234	1234 5th St			
1	External	Mike	555-1234		560		
2	顧客	Joe		777 Oak St		\$5	
2	External	Joe	555-4567		780		
2	ローン	Joseph	555-4567	777 Oak St			\$50,000

\$EA-ID	\$EA-SRC	名前	電話番号	住所	信用リスク	貯金	ローン
3	従業員	Fred	555-9876	543 1st Ave			
4	顧客	Susan		1234 5th St		\$1234	
4	ローン	Sue	555-1234	1234 5th St			\$10,000

エンティティ分析システムでは、外部データ セットの Mike が、共通する電話番号に基づいて、従業員データの Michael と同じであると判断し、彼に ID 1 を割り当てます。

外部データ セットの Joe のケースは、少し難しくなります。顧客の Joe と同一人物なのではないでしょうか?これら 2 つのデータ ソースを区別することはできませんが、Joseph を含む 3 番目のソース、ローンがあります。Joseph の電話番号は、外部データ セットの電話番号と同じです。これに基づき、システムは同一人物であると判断し、ID 2 を割り当てます。

Fred については複数のレコードがないため、ID 3 が割り当てられます。顧客の Susan は、住所が同じであるため、ローンの Sue と同一人物であると見なされ、ID 4 が割り当てられます。

注：これは説明を目的とした最適化した楽観的マッチングの例です。それ自体で単純な名前、電話番号や住所が正確に一致するものではないため、より悲観的なルール セットを選択し、両方のレコードに同じ識別子を割り当てることができます。

## 段階 6 :解決した識別の分析

リポジトリ内の識別の競合を解決した後、今度はその結果についてさらに分析し、処理を実行することができます。たとえば、同じ識別子に重複するデコードが存在する不正なアクティビティの可能性を疑う場合、重複を一覧表示するレポートを作成する必要があります。

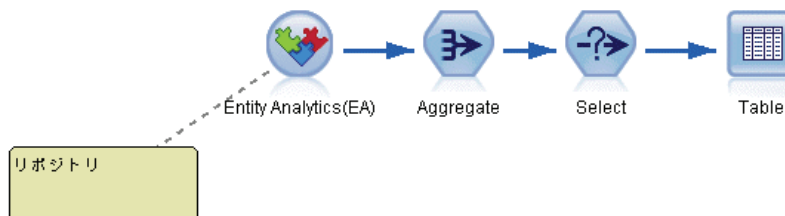
まず、エンティティ分析(EA) 入力ノードを作成し、リポジトリにリンクさせます。

ノードの出力は、以下のフィールドで構成されています。

- システムによって追加された識別子フィールド (段階 5 の例の \$EA-ID)
- システムによって追加されたソースの識別子フィールド (段階 5 の例の \$EA-SRC)
- 段階 4 で指定した一意のキー フィールド

SPSS Modeler で出力を表示するには、表ノードまたはレポート ノードなどの SPSS Modeler 出力ノードを接続し、ストリームのこの部分を実行します。大きくなる場合のある出力を要約する場合、以下の図で示すように、レコード集計ノードやフィールド抽出ノードなどのレコード操作ノードを含めることができます。

図 2-8  
出力の表示



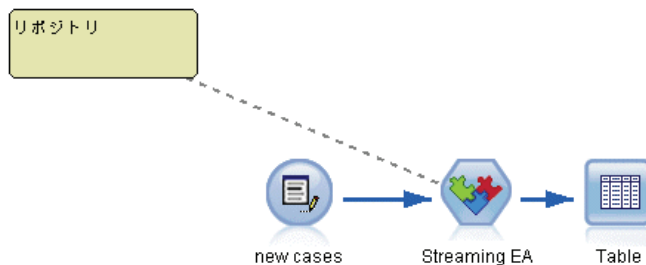
エンティティ分析 (EA) 入力ノードについては、後で詳細に説明します。詳細は、3 章 p.36 解決した識別子を分析するには (Entity Analytics (EA) 入力ノード) を参照してください。

## 段階 7 :レポートリに対して新しいケースを解決する

すべてのデータソース内のすべてのレコードの識別子を解決していますが、新しいレコードのセットを比較して関心のあることとどのような関連があるかを確認する場合、より良いスコアリングを実現するにはどのようにしたらよいのでしょうか？ここで、ストリーミング EA ノードを使用します。

まず、新しい SPSS Modeler データソースのノードを追加し、ストリームに新しいデータを読み込みます。次に、入力ノードをストリーミング EA ノードに接続します。出力を表示するには、前と同じようにテーブル ノードを追加し、ストリームのこの部分が次のように表示されるようにします。

図 2-9  
新しいケースの解決



ストリームのこの部分を実行すると、ストリーミング EA ノードはそれぞれの新しいレコードを読み取り、リポジトリの内容と比較します。リポジトリ内で一致するレコードが見つかった場合、ストリーミング EA ノードは新しいレコードとともに一致するすべてのレコードを出力し、ID およびソース インジケータ フィールドを追加します。一致がない場合、プロセス ノードは ID およびソース インジケータ フィールドが追加された新しいレコードだけを出力します。

これについて説明するために、リポジトリがエンティティ分析 (EA) 入力ノードで出力された内容で構成されていると想定します。 [テーブル 2-1 p. 10](#) を参照してください。

ここで以下の新しいレコードを受け取ります。これらはすでに認識された人物に関連があるでしょうか？

**テーブル 2-2**  
スコアリングされる新しいレコード

名前	住所	電話番号	ローン
Suzan	1234 5th Street	555-1234	\$100,000
Mark	888 9th Ave	555-9999	\$60,000

新しいデータを既存のリポジトリの内容と比較すると、ストリーミング EA ノードでは、既存のレコードの識別子 4 を持つ人物と、最初の新しいレコードが一致します。ただし、2 番目に新しいレコードについては一致はないため、新しい一意の ID 5 が割り当てられます。

ストリーミング EA ノードは識別子とソース インジケータ フィールドを追加し、一致するすべてのレコードとともに新しいレコードを出力します。そのため、出力は以下のようになります。

**テーブル 2-3**  
ストリーミング EA ノードの出力

\$EA-ID	\$EA-SRC	名前	電話番号	住所	信用リスク	貯金	ローン
4	Cust	Susan		1234 5th St		\$1234	
4	ローン	Sue	555-1234	1234 5th St			\$10,000
4	新規ローン	Suzan	555-1234	1234 5th Street			\$100,000
5	新規ローン	Mark	555-9999	888 9th Ave			\$60,000

この出力は、エンティティ分析識別子によって集計され、他の下流ノードに渡してさらに処理を行うことができます。

ストリーミング EA ノードについては、後で詳細に説明します。

## 段階 8 :アラートを生成する

再び、疑わしいアクティビティの可能性が明らかになる場合があります。この場合、ID 4 の人物にはすでに 10,000 ドルのローンがあり、異なる名前で申請を行っていることが考えられます。もちろん、これは承認可能で何ら不正の意思はないと思われませんが、ただし、そのようなアクティビティがビジネス ルールに基づいて疑わしいと判断される場合、調査する必要があります。

たとえば、SPSS Modeler 表ノードまたはレポート ノードを接続および実行して、出力ウィンドウの内容をプリントし、誰かに読ませて手動で警告を行うことができます。また、ストリーミング EA ノードの出力を IBM® SPSS® Modeler で作成されたリスク評価モデルに渡し、より詳細にビジネス ルールを半利するコアのセットを生成します。別の可能性として、エンティティ分析プロセス ノードの出力をデータベースまたは他の媒体にエクスポートし、さらに処理を行います。SPSS Modeler を使用して、特定の要件に対応したアクションを選択できます。

# エンティティ分析のタスク

## タスクについて

ここでは、以下のエンティティ分析タスクについて説明します。

- エンティティ リポジトリの設定
- エンティティ リポジトリの構成
- 解決した識別の分析
- リポジトリに対する新しいケースの解決
- エンティティ リポジトリの消去
- エンティティ リポジトリの削除
- エンティティ分析の管理

## エンティティ リポジトリの設定 (EA エクスポート ノード)

エンティティのリポジトリを設定するプロセスでは、次のタスクが行われます。

1. データ ソースに接続する。詳細は、 p.16 [データ ソースへの接続](#) を参照してください。
2. リポジトリを作成する。詳細は、 p.17 [リポジトリの作成](#) を参照してください。
3. データ ソース内の入力フィールドをリポジトリのフィーチャにマッピングする。詳細は、 p.20 [入力フィールドのフィーチャへのマッピング \(EA エクスポート ノード\)](#) を参照してください。

マッピングを設定したら、現在のデータ ソース、またはリポジトリに読み込まれたすべてのデータ ソースについてマッピングを表示できます。詳細は、 p.24 [フィールド マッピングの表示 \(EA エクスポート ノード\)](#) を参照してください。

## エンティティ リポジトリ

リポジトリは、エンティティ情報のすべてのデータ キャッシュとして機能する集約ストレージ領域を提供します。リポジトリがアクティブであるため、それは単一の状態であり、エンティティ リポジトリによるバージョンングの概念はありません。リポジトリには、すべての入力データ

の現在の状態が保持されているため、リポジトリが非常に大きく成長する場合があります。

使いやすいグラフィカル インターフェイスによりリポジトリの内容を保持できます。 詳細は、 p.25 [エンティティ リポジトリの構成](#) を参照してください。

注： IBM® SPSS® Modeler Premium に付属する IBM® SPSS® Modeler Entity Analytics のバージョンは、IBM SPSS Modeler Entity Analytics とセットの IBM solidDB 製品でホストされている単一リポジトリをサポートします。このバージョンを使用する場合、新しいリポジトリを作成する前に既存のリポジトリを削除する必要があります。IBM SPSS Modeler Entity Analytics への別途ライセンスのアップグレード（IBM SPSS Modeler Entity Analytics Unleashed）を使用して、複数のリポジトリを同一システムに共存させることができます。各リポジトリには 1000 万を超える行を格納し、5 つ以上のプロセッサ コアを使用することができます。詳細については、お近くの IBM サポート担当者にお問い合わせください。

## データ ソースへの接続

入力ノードを使用して、まずソース データを SPSS Modeler に読み込みます。

### データソースへ接続するには

1. SPSS Modeler メイン ウィンドウの最下部にあるノード パレットの [ソース] タブから、ソース データの種類に対応するアイコンをダブルクリックします。アイコンをダブルクリックすると、入力ノードがスクリーン領域に追加されます。
2. スクリーン領域で、アイコンをダブル クリックして、ダイアログ ボックスを表示します。
3. [ファイル] フィールドで、ソース データ ファイルの場所と名前を入力します。
4. 必要に応じてダイアログボックスの残りを入力し（詳細については、[ヘルプ] ボタンをクリック）、[OK] をクリックします。
5. ソース データ ファイルに各レコードを一意に識別するフィールドがない場合、フィールド作成ノードを使用してフィールドを追加します。 詳細は、 p.17 [一意のレコード識別子の追加](#) を参照してください。



## 一意のレコード識別子の追加

エンティティ リポジトリへの入力である各データ ソース ファイルには、各レコードを一意に識別する 1 つのフィールドが必要です。ソース データ ファイルに各レコードを一意に識別するフィールドがない場合、SPSS Modeler フィールド作成ノードを使用してフィールドを追加します。

### 一意のレコード識別子をデータ ソース ファイルに追加するには

1. スクリーン領域で、以前のタスクで追加した入力ノードをクリックします。
2. ノード パレットの [フィールド作成] タブをクリックし、[フィールド生成] アイコンをクリックしてフィールド生成ノードを入力ノードに接続します。
3. スクリーン領域で、フィールド生成ノードをダブル クリックして、ダイアログ ボックスを表示します。
4. [フィールド生成] フィールドで、デフォルト名を追加する識別子フィールドに意味のある名前 (ID など) と置き換えます。
5. [形式を指定して生成] フィールドを [数式] に設定します。
6. [フィールドタイプ] を [連続型] に設定します。
7. [数式] テキスト ボックスで、@INDEX と入力して [OK] をクリックします。

## リポジトリの作成

すべての入力データを格納するリポジトリを作成する必要があります。

### リポジトリを作成するには

1. SPSS Modeler ノードパレットの [エクスポート] タブから、EA エクスポートノードをストリーム領域に配置します。

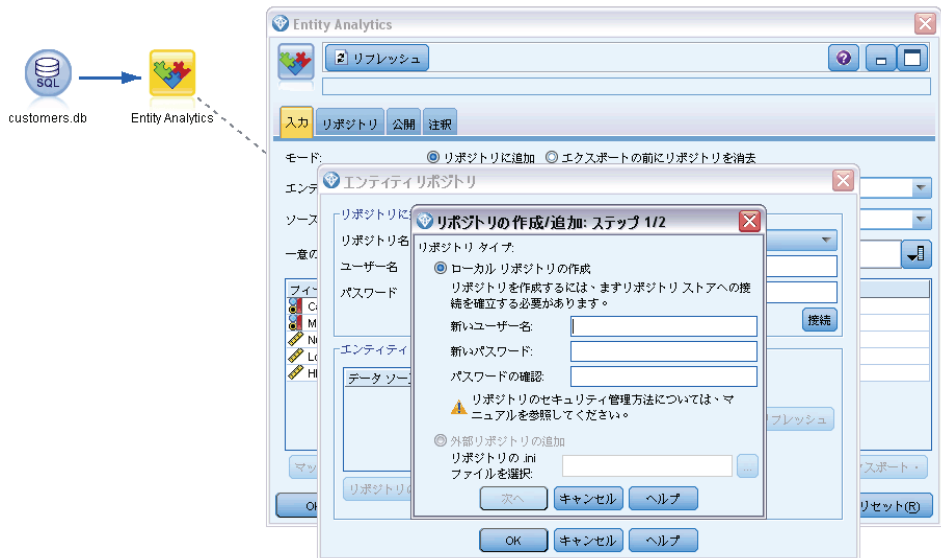
注：初めてリポジトリを作成する場合は、EA エクスポート ノードを使用して、リポジトリ（または一意の識別子フィールドを取得するフィールドを追加した場合は、フィールド生成ノード）に入力するデータを含む SPSS Modeler 入力ノードに接続します。ノードを接続するには、以下の手順を実行します。

- SPSS Modeler 入力ノードを右クリックします。
  - [接続] を選択します。
  - EA エクスポート ノードをクリックします。
2. EA エクスポート ノードをダブル クリックして、ダイアログ ボックスを表示します。

3. [エンティティリポジトリ] リストをクリックします。
4. <[参照...]> をクリックして、[エンティティ解決] ダイアログ ボックスを開きます。
5. [エンティティ解決] ダイアログボックスで、[リポジトリ名] フィールドをクリックします。
6. <[新しいリポジトリを作成...]> をクリックして、リポジトリの作成ウィザードを表示します。

## リポジトリの作成ウィザード

図 3-1  
リポジトリの作成ウィザード



### ステップ 1

ここで、IBM SPSS Modeler Entity Analytics のセットである IBM solidDB 製品を使用してローカル リポジトリを作成するか、リポジトリの外部データベースを使用するかを選択します。

**ローカル リポジトリを作成する:** 作成しているリポジトリをホストしている IBM solidDB データベースの管理者ユーザー名およびパスワードを指定します。パスワードを確認して [次へ] をクリックします。

注：後で管理者の資格情報を変更する必要がある場合、データベースのコマンドライン エディタを使用します。詳細は、p.46 [リポジトリ データベースの管理者の資格情報の管理](#) を参照してください。

**外部リポジトリを追加:** リポジトリをホストする外部データベースを使用したい場合にこのオプションを使用します。[リポジトリの .ini ファイルを選択] フィールドでデータベース .ini ファイルの場所を入力し、[次へ] をクリックします。

## ステップ 2

**新しいリポジトリ名:** 新しいリポジトリの一意の名前を入力します。

**設定のインポート元:** (ローカル リポジトリのみ) 既存のリポジトリのに基づいて設定する場合、リポジトリを選択し、そうでない場合は [デフォルト] を選択します。詳細は、 p.25 エンティティ リポジトリの構成 を参照してください。

既存のリポジトリを選択した場合、それらが前の画面で入力したものと異なるときは、接続の詳細を入力してください。

[OK] をクリックして、新しいリポジトリを作成し、[エンティティ解決インスタンス] ダイアログボックスを表示し、そこでリポジトリに接続できます。

## エンティティ リポジトリのオプション

[エンティティ解決] ダイアログボックスには、エンティティ リポジトリの作成、接続、設定、および維持するためのオプションがあります。

図 3-2  
エンティティ リポジトリのオプション



**リポジトリへの接続:** これらのオプションを使用して、新しいエンティティのリポジトリを作成するか、または既存のものに接続します。

- **リポジトリ名:** 存在する場合、現在のエンティティのリポジトリを示しています。複数存在する場合に別のリポジトリを選択するには、リストからいずれかを選択します。

新しいリポジトリを作成する場合、[<新しいリポジトリを作成...>] をクリックします。これによりウィザードが起動し、作成プロセスを指示します。

- **ユーザー名:** 選択したリポジトリのユーザー名を入力します。
- **パスワード:** ユーザー名に対するパスワード。
- **接続:** 現在のリポジトリに接続します。

**エンティティリポジトリの管理:** 表には現在のリポジトリ（接続しているもの）に読み込まれているデータソースが表示され、各データソース内のレコード数を示します。

- **リフレッシュ:** 例えば、新しいデータソースを追加するか、既存のデータソースのサイズを変更した場合、テーブル内のデータソースおよびサイズ情報を更新します。
- **リポジトリの消去:** リポジトリからすべてのソースデータを削除しますが、すべての構成の詳細を保持します。設定情報が有用であるにもかかわらず、リポジトリからすべてのデータレコードを削除する場合はこのオプションを使用できます。詳細は、p. 48 [エンティティリポジトリの消去](#) を参照してください。

**リポジトリ全体の削除:** 現在のリポジトリの内容と設定の詳細を完全に削除します。詳細は、p. 49 [エンティティリポジトリの削除](#) を参照してください。

**リポジトリの設定:** ウィンドウで現在のリポジトリを設定で行きます。詳細は、p. 25 [エンティティリポジトリの構成](#) を参照してください。

## 入力フィールドのフィーチャへのマッピング (EA エクスポート ノード)

リポジトリには、標準として多くの機能があります。異なるデータソースでは、同じフィーチャに対応する情報タイプについて、それぞれ別のフィールド名を使用します (例: Address1 または Address Line 1) for information types that correspond to the same feature. 重複を避けるために、入力データソースフィールドを特定のリポジトリフィーチャにマッピングする必要があります。データセットのすべてのフィールドをマッピングする必要はなく、他のデータセットの同じフィーチャに対応する可能性があるフィールドだけをマッピングします。

データ ソースがリポジトリ内の事前定義されていない他のタイプの情報に対応するフィールドを使用している場合、[リポジトリ設定] ウィンドウから新しいフィーチャを作成することができます。詳細は、p. 25 [エンティティ リポジトリの構成](#) を参照してください。

### 入力フィールドをフィーチャにマッピングするには

1. EA エクスポート ノードをストリーム領域のデータ ソース ノードへに接続します。使用する各データ ソース ノードは独自の EA エクスポート ノードに接続する必要があります。
2. EA エクスポート ノードを開いて、入力フィールドをマッピングするオプションが用意された [入力] タブを表示します。詳細は、p. 21 [マッピングするリポジトリ入力オプション](#) を参照してください。
3. EA エクスポート ノードで、[リポジトリ] タブを選択し、現在のデータ ソース、または複数のデータ ソースを使用している場合はすべてのデータ ソースのマッピングの割り当てを表示します。
4. 一連のマッピングの割り当てを保存するには（別のデータ ソース ノードで使用する場合など）、[エクスポート マッピング] をクリックします。

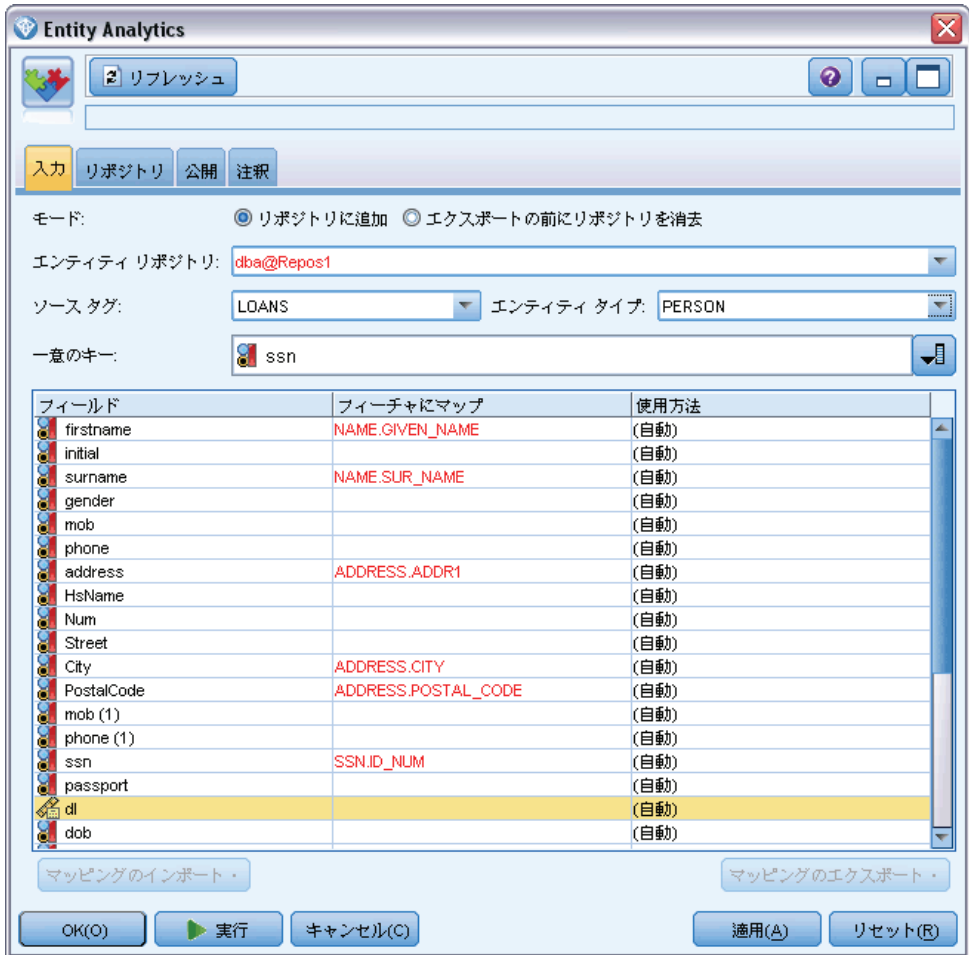
最初のデータ ソース ノードのマッピングが終了したら、使用する他のデータ ソース ノードのプロセスを繰り返します。

### マッピングするリポジトリ入力オプション

[入力] タブには、リポジトリへのエクスポートの準備が整ったリポジトリのフィーチャにデータ ソースのフィールドをマッピングするためのオプションが含まれています。このタブでマッピングの割り当てを設定し、[リポジトリ] タブをクリックして他のデータ ソースのマッピングを表示し、[実行] をクリックしてデータをリポジトリにエクスポートします。

一連のマッピングをすでに XML ファイルに保存している場合、[マッピングのエクスポート] をクリックして使用することができます。

図 3-3  
入力フィールドのフィーチャへのマッピング



**モード:** リポジトリの既存コンテンツにソース ファイルのレコードを追加する場合、デフォルトの選択 [リポジトリへの追加] のままにします。ソースレコードを追加する前にリポジトリ コンテンツをクリアするが、設定情報を維持したい場合、[エクスポート前にリポジトリを消去] を選択します。

**エンティティ リポジトリ:** 存在する場合、現在のエンティティのリポジトリを示しています。複数存在する場合に別のリポジトリを選択するには、リストからいずれかを選択します。新しいリポジトリを作成するには、[<参照...>] を選択してダイアログ ボックスを表示し、リポジトリを作成します。詳細は、p. 19 エンティティ リポジトリのオプション を参照してください。

**ソース タグ:** 現在リポジトリに認識されているデータソースを示すタグのリスト。リストからソース タグを選択するか、[<新しいソース タグを追加...>] を選択して新しいデータ ソースのタグを作成します。

**エンティティタイプ:** リポジトリに定義されるエンティティ タイプのリスト (機能のセット)。リストからエンティティ タイプを選ぶか、[<新しいエンティティタイプを追加...>] をクリックして [リポジトリ設定] ウィンドウを表示し、新しいエンティティ タイプを定義します。詳細は、p. 25 [エンティティ リポジトリの構成](#) を参照してください。

**一意のキー:** (必須) データ レコードの一意の識別子に使用する入力フィールド。

**マッピング テーブル:** このテーブルでは、リポジトリ内の対応する機能への各入力フィールドをマッピングすることができます。適切な機能が選択したエンティティタイプに存在しない場合は、ここで新しい機能を作成することができます。

- **フィールド:** 選択したデータ ソース内の入力フィールドのセット。各フィールドには、フィールドの尺度 (データ型) を示すアイコンがあります。
- **機能にマップ:** フィールドを機能にマッピングするには、フィールドの行でこの列をダブルクリック (またはスペースバーを押す) し、リストから機能を選択します。適切な機能を使用できない場合、[<新しいエンティティタイプを追加...>] をクリックして [リポジトリ設定] ウィンドウを表示し、新しいエンティティ タイプを定義します。詳細は、p. 25 [エンティティ リポジトリの構成](#) を参照してください。
- **使用方法:** 自宅と職場の電話番号など、複数のコンテキストが可能な特定のフィールドのコンテキストを示します。ADDRESS および PHONE フィーチャに使用可能なプリセットの使用方法があり、すべてのフィーチャに独自の使用方法を作成できます。デフォルトの (自動) 以外の使用方法を設定する場合、設定する行のこの列をクリックして、既存の使用方法があれば選択するか、[<使用方法を追加...>] をクリックして新しい使用方法を作成します。詳細は、p. 31 [エンティティ タイプの管理](#) を参照してください。

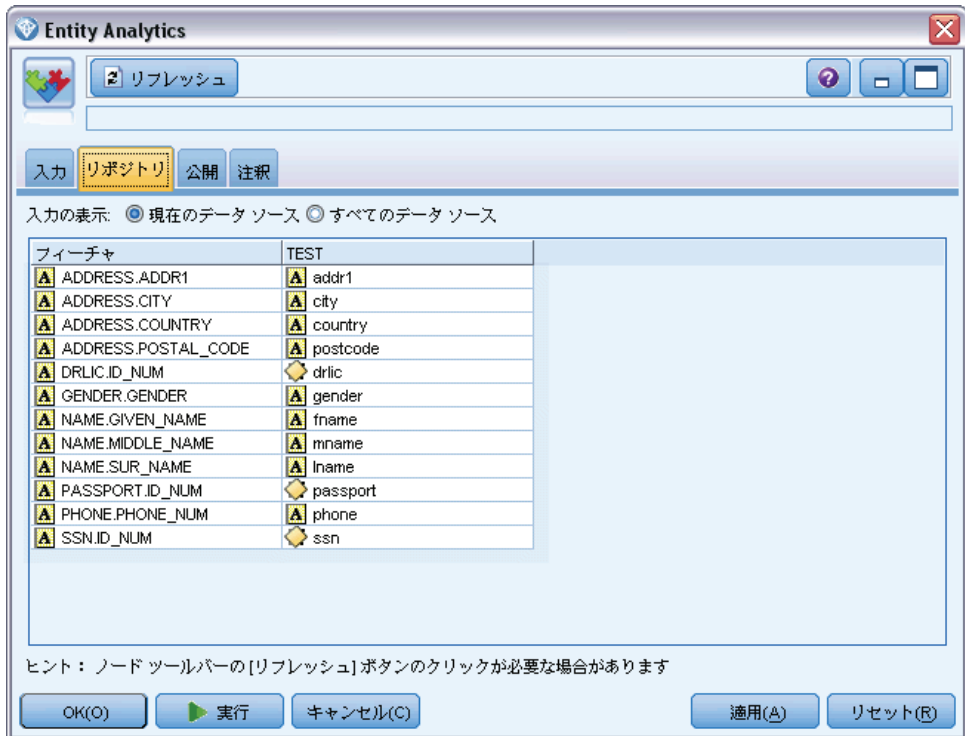
**マッピングのインポート:** 以前エクスポートされたフィールドからフィーチャへのマッピングを外部 XML ファイルからインポートします。それは、異なるソースに対して同じマッピングを再定義する必要がなくなるため、同じマッピング要件を持つさまざまなデータソースがある場合に便利です。

**マッピングのエクスポート:** マッピング テーブルに表示されたフィールドからフィーチャへのマッピングのセットを外部 XML ファイルにエクスポートします。

## フィールド マッピングの表示 (EA エクスポート ノード)

[リポジトリ] タブで、[更新] ボタンをクリックし、入力フィールドがマッピングされたリポジトリ フィーチャを確認します。現在のデータソース（このエクスポートのノードに接続されたソースノードによって制御）、またはすべてのデータソースのいずれかのフィーチャを確認できます。

図 3-4  
フィールド マッピングの表示



**入力を表示：**現在のデータ ソース、またはリポジトリに読み込まれたすべてのデータ ソースについてマッピングを表示するオプションを選択します。

**リフレッシュ：**選択されたオプションの表示を更新します。

**フィーチャ。**表示されたデータ ソースにマッピングがあるすべてのフィーチャのリスト。マッピングされていないフィーチャは表示されません。

**<データソース>**各列には、マッピングが定義されている各フィーチャの特定のデータ ソースのマップされたフィールドが表示されます。



## エンティティリポジトリの構成

リポジトリ全体に視覚的な使いやすいインタフェースを提供する [リポジトリの構成] ウィンドウから、リポジトリの内容を変更することができます。

同じまたは類似した構成の複数のリポジトリを使用する場合、基本的な構成を設定し、他のリポジトリにインポートできるファイルにエクスポートできます。詳細は、p. 35 [リポジトリ設定の再利用](#) を参照してください。

注意:すでにデータを含むリポジトリの構成を変更および保存する場合、リポジトリ コンテンツを消去してデータをリロードすることをお勧めします。そうすることにより、リポジトリが一貫しない状態になることを回避することができます。

### リポジトリ構成を設定するには

1. エンティティ分析ノードを開きます。
2. [エンティティリポジトリ] リストをクリックします。
3. <[参照...]> をクリックして、[エンティティ解決インスタンス] ダイアログ ボックスを開きます。
4. [エンティティ解決インスタンス] ダイアログボックスで、[リポジトリ名] リストをクリックします。
5. 構成を設定するリポジトリを選択します。
6. 接続されていない場合、管理者のユーザー名とパスワードを入力し、[接続] をクリックします。
7. [リポジトリの設定] ボタンが有効な場合、クリックすると [リポジトリの設定] ウィンドウが表示されます。
8. 以下のセクションで説明されているように構成の詳細を作成します。

[リポジトリの設定] ウィンドウの左側のナビゲーション ペインには、リポジトリの別の特性を管理できるツリー構造が表示されます。

テーブル 3-1  
[リポジトリの設定] ウィンドウの主要な要素

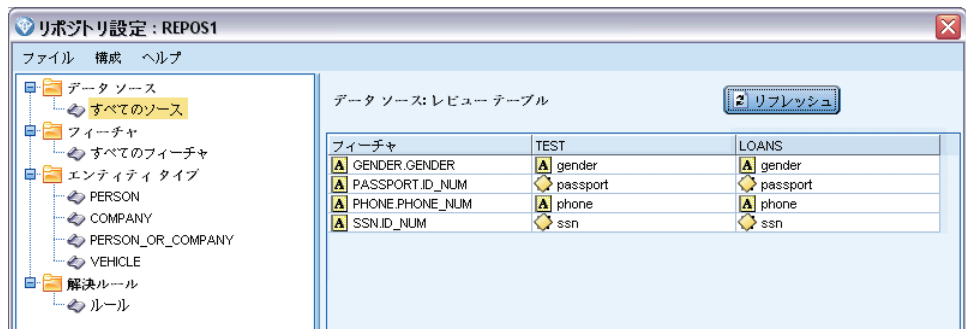
セクション	説明	
データ ソース	さまざまなリポジトリ フィーチャに対するすべてのデータ ソースのマッピングを表示します。	詳細は、p. 26 <a href="#">データ ソース マッピングの表示</a> を参照してください。
フィーチャ	新しいフィーチャを作成、または既存のフィーチャを複製、編集したり削除します。	詳細は、p. 26 <a href="#">リポジトリのフィーチャの更新</a> を参照してください。

セクション	説明	
エンティティ タイプ	新しいエンティティ タイプを作成、または既存のエンティティタイプを管理（複製、名前変更、フィーチャの接続または削除、削除）します。	詳細は、 p.31 <a href="#">エンティティ タイプの管理</a> を参照してください。
解決ルール	エンティティの一致に対ししきい値を設定します。	詳細は、 p.34 <a href="#">エンティティの一致に対するしきい値の設定</a> を参照してください。

## データ ソース マッピングの表示

[リポジトリの設定] ウィンドウの [データ ソース] セクションで、[すべてのソース] エントリは、さまざまなリポジトリ機能へのすべてのデータソースのマッピングを読み取り専用で表示します。

図 3-5  
設定ウィンドウのデータソース マッピング



新しいデータ ソースがリポジトリに追加された場合、[更新] をクリックしてリストを更新します。

注：ここでリポジトリにデータ ソースを追加することはできません。データソースは、SPSS Modeler 入力ノードを作成し、Entity Analytics エクスポート ノードに接続することによって追加することができます。詳細は、 p.16 [データ ソースへの接続](#) を参照してください。

## リポジトリのフィーチャの更新

リポジトリ フィーチャは、エンティティ データ ソースで使用できる個別の情報タイプです。特定のデータソース固有のフィーチャ、（姓、名、生年月日など）さまざまなデータ ソースで使用されることができ、フィーチャには 1 つ以上の要素を含むことができ

ます。各要素は通常、データ レコードのフィールド、またはデータベース テーブルの列に対応します。

[リポジトリの設定] ウィンドウの [フィーチャ] セクションで、[すべてのフィーチャ] エントリはすべてのリポジトリのフィーチャを管理することができます。以下の操作が可能です。

- 新規フィーチャの作成
- 既存フィーチャの複製（たとえば、既存フィーチャに基づいて新しいフィーチャを作成）
- 既存のフィーチャを編集
- 既存のフィーチャを削除

これらのタスクについては、このセクションの後半で説明します。

図 3-6  
設定ウィンドウに表示されたフィーチャ



フィーチャ リストには、このリポジトリで定義されているすべてのフィーチャを示しています。リストの列は、フィーチャが持つさまざまなプロパティを表示します。

**フィーチャ:** フィーチャの名前フィーチャ名の隣の南京錠のシンボルは、フィーチャがロックされていることを示します。ロックされているフィーチャは削除することも複製することもできません。変更を保存することもできません。

**度数:** このフィーチャの同じ値を持つことができる識別子の数を示します。値は、なし、名前、1（例： パスポート番号）、少数（例： 住所）、多数（例： 誕生日）です。

**排他性:** エンティティは通常この種類のフィーチャだけであることを示します。たとえば、誕生日や身分証番号などは値が [はい] となりますが、住所やクレジット カード番号の値は [いいえ] となります（エンティティは複数の住所またはクレジット カードとなる場合があるため）。

**安定性:** このフィーチャの安定性の値（エンティティの寿命の間変わる可能性が低いかどうか）を示します。たとえば、誕生日フィーチャは変更されないため [はい] という値が割り当てられます。一方住所は変更の可能性が高く、安定性が低いため [いいえ] となります。注：性別は通常生涯を通じて一貫していますが、データの誤りにより度数が不正に指定されている場合があるため。デフォルトの設定は値が [いいえ] となっています。

#### 新しいフィーチャを作成するには

1. 次のいずれかを実行します。
  - [新しいフィーチャの作成] ボタン（画面の右上の一番上のボタン）をクリックします。
  - 画面左側のナビゲーション ペインの [すべてのフィーチャ] を右クリックし、[新しいフィーチャ] を選択します。
2. [フィーチャの追加/編集] ダイアログボックスを入力します。詳細は、p. 29 [フィーチャの追加または編集](#) を参照してください。

#### 既存のフィーチャを複製するには

1. 画面の右側のテーブルの [フィーチャ] 列で、複製するフィーチャを選択します。
2. [選択したフィーチャの複製] ボタン（画面の右上の 2 番目のボタン）をクリックします。
3. [フィーチャの追加/編集] ダイアログボックスを入力します。詳細は、p. 29 [フィーチャの追加または編集](#) を参照してください。

#### 既存のフィーチャを編集するには

注意: リポジトリにデータが含まれているときにフィーチャまたはフィーチャ要素を編集または削除する場合、リポジトリを消去してデータをリロードする必要があります。そうすることにより、リポジトリが一貫しない状態になることを回避することができます。

1. 画面の右側のテーブルの [フィーチャ] 列で、編集するフィーチャを選択します。注：作成したフィーチャのみ編集が可能であり、システムが提供するフィーチャは編集できません。
2. [選択したフィーチャの編集] ボタン（画面の右上の 3 番目のボタン）をクリックします。
3. [フィーチャの追加/編集] ダイアログボックスを入力します。詳細は、p. 29 [フィーチャの追加または編集](#) を参照してください。

### 既存のフィーチャを削除するには

注意:リポジトリにデータが含まれているときにフィーチャまたはフィーチャ要素を編集または削除する場合、リポジトリを消去してデータをリロードする必要があります。そうすることにより、リポジトリが一貫しない状態になることを回避することができます。

1. 画面の右側のテーブルの [フィーチャ] 列で、削除するフィーチャを選択します。注：作成したフィーチャのみ削除が可能であり、システムが提供するフィーチャは削除できません。
2. 次のいずれかを実行します。
  - [選択したフィーチャの削除] ボタン（画面の右の一番下のボタン）をクリックします。
  - 画面左側のナビゲーション ペインの [すべてのフィーチャ] を右クリックし、[削除] を選択します。
3. [続行] をクリックしてフィーチャの削除を確定します。注：フィーチャの削除を取り消すことはできません。

### フィーチャの追加または編集

注意:リポジトリにデータが含まれているときにフィーチャまたはフィーチャ要素を編集または削除する場合、リポジトリを消去してデータをリロードする必要があります。そうすることにより、リポジトリが一貫しない状態になることを回避することができます。

[フィーチャの追加/編集] ダイアログ ボックスで、新しいリポジトリフィーチャを作成、または既存のフィーチャを複製または編集することができます。

図 3-7  
フィーチャの編集

.	エレメント	説明	データ型	検索	比較	説明
1	ADDR_TY...	Address t...	文字列	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
2	FULL_ADDR	Full address	文字列	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3	ADDR1	Addr line 1	文字列	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
4	ADDR2	Addr line 2	文字列	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
5	ADDR3	Addr line 3	文字列	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
6	CITY	City	文字列	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
7	STATE	State	文字列	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
8	POSTAL_...	Postal code	文字列	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
9	COUNTRY	Country	文字列	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

**フィーチャの種類:** フィーチャが関連する情報のタイプを示すラベル。このラベルはフィーチャの識別子の最初の部分です。

**説明:** フィーチャの種類に関する簡単なテキストの説明。

**度数:** このフィーチャの同じ値を持つことができる識別子の数を示します。値は、なし、名前、1 (例: パスポート番号)、少数 (例: 住所)、多数 (例: 誕生日) です。

**排他性:** エンティティは通常この種類のフィーチャだけであることを示します。たとえば、誕生日や身分証番号などは値が [はい] となりますが、住所やクレジット カード番号の値は [いいえ] となります (エンティティは複数の住所またはクレジット カードとなる場合があるため)。

**安定性:** このフィーチャの安定性の値 (エンティティの寿命の間変わる可能性が低いかどうか) を示します。たとえば、誕生日フィーチャは変更されないため [はい] という値が割り当てられます。一方住所は変更の可能性が高く、安定性が低いため [いいえ] となります。注: 性別は通常生涯を通じて一貫していますが、データの誤りにより度数が不正に指定されている場合があるため。デフォルトの設定は値が [いいえ] となっています。

**履歴を保持する:** [はい] に設定 (デフォルト) すると、変更履歴をフィールド値に保持します。これは、顧客の住所フィールドなどに役立ちます。この場合、顧客が住所を数回変更した場合、変更履歴を保持すると一致率が

改善します。ただし、勘定残高等のフィールドの場合、[いいえ] に設定して、過剰なデータ量を保持しないようにする必要があります。

**要素テーブル:** このフィーチャを構成する要素のリスト。

- **要素:** 要素名。
- **説明:** 要素が提供する簡単な説明文。
- **データ型:** この要素で使用できるデータの種類。
- **検索:** このオプションを選択すると、フィーチャのこの要素をそれ自体だけで、またはフィーチャ ラベル検索のその他の要素と組み合わせて検索することができます。たとえば、PASSPORT フィーチャに ID\_NUM と COUNTRY の 2 つの要素があるとします。ID\_NUM だけについて検索できるようにしたい場合、ID\_NUM の [検索] ボックスを選択して COUNTRY の [検索] ボックスをクリアします。国や州など、また発行日や有効期限日など、フィーチャに対する変更を検索の一部とすることはできません。
- **比較:** エンティティ解決の比較ルーティンに送信するフィーチャの要素を示します。ID\_NUM および COUNTRY の 2 つの要素を持つ PASSPORT フィーチャについて、COUNTRY は常に入力されるとは限らないため、ID\_NUM だけを比較ルーティンに送信する必要があります。
- **説明:** その内部の説明に表示されるフィーチャの要素を示します。ID\_NUM および COUNTRY PASSPORT の 2 つの要素を持つ PASSPORT の例で、フィーチャの説明に療法の要素を表示したい場合、2 つの要素の [説明] ボックスを選択します。ただし、フィーチャに追加の要素 ISSUE\_DT があるとします。[説明] ボックスをクリアして、フィーチャの説明に表示されないようにします。

**新しい要素の追加:** 新しい要素を定義できるよう、新しい行を要素テーブルに追加します。

**要素の削除:** 選択した行を要素テーブルから削除します。この操作は取り消しできません。

注意: リポジトリにデータが含まれているときにフィーチャまたはフィーチャ要素を編集または削除する場合、リポジトリを消去してデータをリロードする必要があります。そうすることにより、リポジトリが一貫しない状態になることを回避することができます。

## エンティティタイプの管理

**エンティティ タイプ**は論理的に属するリポジトリ フィーチャの指定のセットです。たとえば、顧客データ セットで使用することを目的としたエンティティ タイプは、名前、生年月日、性別、住所、電話番号などのフィーチャから成っています。

IBM SPSS Modeler Entity Analytics リポジトリはエンティティ タイプの標準セットに付属し、独自のエンティティ タイプを追加できます。

[リポジトリの設定] ウィンドウの [エンティティ タイプ] セクションは、作成されているさまざまなエンティティのタイプを示します。以下の操作が可能です。

- エンティティ タイプを作成する
- 既存エンティティ タイプを複製する（たとえば、既存エンティティ タイプに基づいて新しいエンティティ タイプを作成）
- フィーチャをエンティティ タイプに接続する
- フィーチャをエンティティ タイプから削除する
- エンティティ タイプの名前を変更する
- エンティティ タイプを削除する

図 3-8  
エンティティタイプに属するフィーチャの表示



**エンティティタイプ:** 選択したエンティティ タイプの名前

**フィーチャ:** このエンティティ タイプを構成する有効なフィーチャのリスト。

**使用タイプ:** (オプション) この機能を使用する別のコンテキストを示しています。この列をダブルクリックして使用タイプを追加または編集し、カンマとスペースでの使用タイプを区切ります。ここで指定した値は、ユーザーが [入力] タブのフィーチャの [使用方法] 列をクリックした場合に EA エクスポートまたはストリーミング EA ノードに表示される値を定義します。詳細は、p. 21 [マッピングするリポジトリ入力オプション](#) を参照してください。

**エンティティタイプを新規作成するには**

1. 画面左側のナビゲーション ペインの [エンティティタイプ] を右クリックします。



2. [新しいエンティティタイプ] を選択します。
3. エンティティ タイプの一意の名前を入力して、[OK] をクリックします。
4. フィーチャをエンティティ タイプに接続します (次のセクションを参照)。

#### **フィーチャをエンティティ タイプに接続するには**

1. 画面左側のナビゲーション ペインでエンティティ タイプを選択します。
2. [フィーチャの接続] ボタン (画面の右の一番上のボタン) をクリックします。
3. 使用可能なフィーチャのリストから、フィーチャを選択 (複数のフィーチャを選択する場合は Ctrl + クリック) して [OK] をクリックします。

#### **フィーチャをエンティティ タイプから削除するには**

1. 画面左側のナビゲーション ペインでエンティティ タイプを選択します。
2. 画面右の接続されたフィーチャのテーブルから 1 つまたは複数のフィーチャを選択します。複数のフィーチャを選択するには Ctrl を押しながらクリックします。
3. [フィーチャの接続解除] ボタン (画面の右の一番下のボタン) をクリックします。

#### **既存のエンティティ タイプを複製するには**

1. 画面左側のナビゲーション ペインで、複製するエンティティ タイプを右クリックします。
2. [エンティティタイプの複製] を選択します。
3. 新しいエンティティ タイプの一意の名前を入力して、[OK] をクリックします。
4. 必要に応じて、フィーチャをエンティティ タイプに接続したり、エンティティ タイプから削除します (前の説明を参照)。

#### **エンティティ タイプの名前を変更するには**

注意: リポジトリにデータが含まれているときにフィーチャまたはフィーチャ要素を編集または削除する場合、リポジトリを消去してデータをリロードする必要があります。そうすることにより、リポジトリが一貫しない状態になることを回避することができます。

1. 画面左側のナビゲーション ペインで、名前を変更するエンティティ タイプを右クリックします。

2. [名前の変更] を選択します。
3. エンティティ タイプの新しい名前を入力して、[OK] をクリックします。

### エンティティ タイプを削除するには

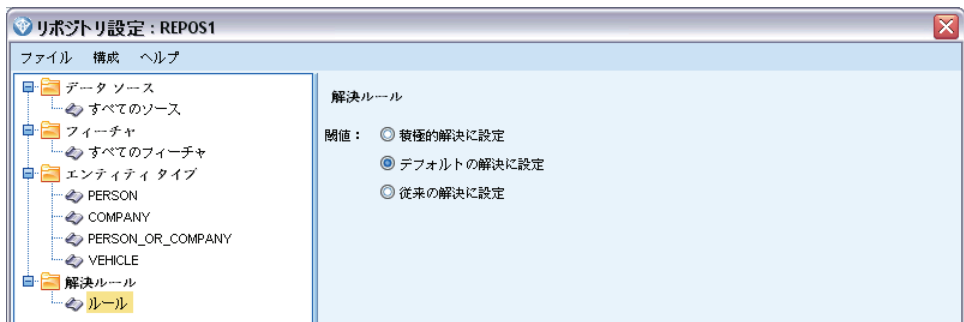
注意: リポジトリにデータが含まれているときにフィーチャまたはフィーチャ要素を編集または削除する場合、リポジトリを消去してデータをリロードする必要があります。そうすることにより、リポジトリが一貫しない状態になることを回避することができます。

1. 画面左側のナビゲーション ペインで、削除するエンティティ タイプを右クリックします。
2. [削除] を選択します。
3. [OK] をクリックしてエンティティ タイプの削除を確定します。注意: エンティティ タイプの削除を取り消すことはできません。

## エンティティの一致に対するしきい値の設定

[リポジトリの設定] ウィンドウの [解決ルール] セクションで、エンティティの一致が発生するしきい値を選択します。

図 3-9  
エンティティ解決ルールの設定



リポジトリを作成すると、一致はデフォルトのしきい値にプリセットされます。

エンティティ解決を実行するための十分な一致が記録にない場合、[積極的な解決に設定] を選択します。

他の設定からデフォルトのしきい値に戻す場合は、[デフォルトの解決に設定] をクリックします。

一致が多すぎる場合、[従来 of 解決に設定] を選択します。

## リポジトリ設定の再利用

すでに設定しており別のリポジトリに使用する場合、XML ファイルに既存の設定をエクスポートし、他の（ターゲット）にリポジトリにファイルをインポートすることができます。

### 既存の設定を再利用するには

1. 設定を使用するリポジトリの [リポジトリの設定] ウィンドウを表示します。詳細は、p. 25 [エンティティ リポジトリの構成](#) を参照してください。
2. メニューから次の項目を選択します。  
設定 > 設定のエクスポート
3. [名前を付けて保存] ダイアログ ボックスで、エクスポート XML ファイルの名前と場所を選択します。
4. 対象リポジトリの [リポジトリの設定] ウィンドウを表示します。
5. メニューから次の項目を選択します。  
設定 > 設定のインポート
6. [開く] ダイアログ ボックスで、以前エクスポートされた XML ファイルの名前と場所を選択し、[開く] をクリックします。

## 設定の変更の保存

### 設定の変更を保存するには

[リポジトリの設定] ウィンドウのメニューから、次の項目を選択します。  
ファイル > 保存

## 設定ウィンドウを閉じる

### 設定ウィンドウを終了するには

[リポジトリの設定] ウィンドウのメニューから、次の項目を選択します。  
ファイル > 終了

設定の変更を保存していない場合、変更を保存して終了する場合は [OK] を、保存しないで終了する場合には [キャンセル] をクリックします。

## 解決した識別子を分析するには (Entity Analytics (EA) 入力ノード)

少なくとも 1 つのデータ ソースをリポジトリに入力した場合、Entity Analytics (EA) 入力ノードを使用して解決した識別子を他の IBM® SPSS® Modeler ノードに渡し、解決済み識別子を表示するレポートを作成するなど、さらなる分析または処理を行うことができます。

### 解決した識別を分析するには

1. Entity Analytics (EA) 入力ノードをストリームに追加します。
2. Entity Analytics (EA) 入力ノードを開きます。
3. [データ]タブで、エンティティのリポジトリと入力データソースを選択します ([リフレッシュ] をクリックしてレコード数を更新します)。詳細は、p. 36 データ ソースの選択 を参照してください。
4. さらにノードをストリームに追加し、必要に応じて処理を実行します。詳細は、p. 38 ノードのストリームへの追加 を参照してください。

## データ ソースの選択

[データ]タブで、さらに処理を実行するリポジトリ内の 1 つ以上のデータ ソースを選択します。データ ソースのレコード数を更新するには、[リフレッシュ] をクリックします。

図 3-10  
入力ノードのデータ ソースを選択



**エンティティ リポジトリ:** 存在する場合、現在のエンティティのリポジトリを示しています。複数存在する場合に別のリポジトリを選択するには、リストからいずれかを選択します。新しいリポジトリを作成するには、[<参照...>] を選択してダイアログ ボックスを表示し、リポジトリを作成します。詳細は、p. 19 エンティティ リポジトリのオプション を参照してください。

**ソースのレコードを含める:** この表には、各ソース内のレコード数とともに、リポジトリに入力されたさまざまなデータ ソースが示されています。さらなる分析および処理を実行するために使用するデータ ソースの [含める] チェックボックスを選択します。

## データ フィールドの名前の変更

[フィルタ] タブを使用して、さらなる処理のために下流に渡された解決した ID フィールドのいずれかの名前を変更することができます。たとえば、別のデータセットを使用して下流を結合する場合にフィールド名の互換性を維持するなど、解決の ID フィールドの名前を変更したい場合があります。

元の名前のフィールドは次のとおりです。

テーブル 3-2  
解決した識別子のフィールド

フィールド	説明
\$EA-ID	エンティティ識別子
\$EA-SRC	レコードが作成元であるデータ ソースを示すソース タグ
\$EA-KEY	データ ソース ファイルの一意キーとして指定されているフィールド

注： また、[フィルタ] タブを使用してフィールドを除外することができますが、解決した ID フィールドはエンティティの分析処理に最低限必要であるため、ここで行うべきではありません。

## データ フィールドのデータ型情報の設定

[データ型] タブで、さらなる処理のために下流に渡される解決した ID フィールドのさまざまなプロパティを表示したり変更したりできます。

変更できるプロパティは、通常の SPSS Modeler データ型ノードの [データ型] タブのプロパティと同じです。プロパティは次のとおりです。

テーブル 3-3  
フィールドのデータ型プロパティ

プロパティ	説明
尺度	フィールドのデータの特性を説明する尺度 (つまりデータ型)。
値	データセットからデータ値を読み込むオプションを指定します。
欠損値	フィールドの欠損値の処理方法を指定します。

プロパティ	説明
検査	フィールドの値が指定された値または範囲に適合させるための検証オプション。
役割	データ モデリング ノードまたはモデル ナゲットに渡された場合、このフィールドが使用される方法を指定します。

## ノードのストリームへの追加

さまざまな SPSS Modeler ノードをストリームに追加して、Entity Analytics (EA) 入力ノードの出力に分析または処理を実行します。たとえば、次のいずれかまたは複数を追加することができます。

- 非常に大きくなる場合のある出力を要約するレコード集計ノード
- 出力のサブセットを選択する条件抽出ノード
- Entity Analytics (EA) 入力ノードの出力を表示するテーブル ノード
- レポートの出力を印刷するレポート ノード
- スプレッドシートまたはデータベースなど、出力を異なる形式にエクスポートする SPSS Modeler エクスポート ノード

詳細については、『IBM SPSS Modeler 入力、処理、出力ノード ガイド』の「レコード設定、出力、エクスポート ノード」に関するセクションを参照してください。

## リポジトリに対して新しいケースを解決する（ストリーミング EA ノード）

すでにリポジトリにいくつかのアイデンティティの解決を実行した場合、その後出現した新しいケースと比較するためにストリーミング EA ノードを使用することができます。このノードが新しいデータソースのレコードを処理し、すでにリポジトリ内の解決したエンティティとそれらを比較して、さらなる処理のために一致するすべてのレコードを渡します。一致は正確に、またはより緩く、既存のエンティティに関連するように設定することができます。

EA エクスポート ノードと同様、ストリーミング EA ノードは入力として単一の SPSS Modeler 入力ノードを採用します。ただし、ストリーミング EA ノードは次の点で異なります。その入力レコードに関連するすべてのエンティティについてエクスポート ノードはレコードを出力しますが、ストリーミング EA ノードは、すでにリポジトリに解決のエンティティに関連するものだけエンティティに関するレコードを出力します。詳細は、p. 44 [ストリーミング EA ノードの出力](#) を参照してください。

### リポジトリに対して新しいケースを比較するには

1. 既存のエンティティと比較する新しいレコードを含むデータソースに接続します。詳細は、p.16 [データソースへの接続](#) を参照してください。
2. [レコード設定] タブで、ストリーミング EA ノードをデータソースノードに接続します。
3. Entity Analytics エクスポート ノードをダブルクリックして、ダイアログボックスを表示します。
4. [エンティティリポジトリ] リストをクリックします。
5. <[参照...]> をクリックして、[エンティティ解決] ダイアログボックスを開きます。
6. [エンティティ解決] ダイアログボックスで、[リポジトリ名] フィールドをクリックします。
7. 使用したいリポジトリの名前をクリックします。
8. リポジトリのユーザー名とパスワードを入力し、[接続] をクリックします。リポジトリに接続したら、[OK] をクリックします。
9. [ストリーミング EA] ダイアログボックスで、使用するエンティティタイプを選択します。詳細は、p.31 [エンティティタイプの管理](#) を参照してください。
10. データソース内の入力フィールドをリポジトリのフィーチャにマッピングする。詳細は、p.40 [入力フィールドのフィーチャへのマッピング \(ストリーミング EA ノード\)](#) を参照してください。
11. オプションで [リポジトリ] タブをクリックし、このデータソースのマッピング (またはマッピングされているすべてのデータソース) を示すテーブルを表示します。詳細は、p.41 [フィールドマッピングの表示 \(ストリーミング EA ノード\)](#) を参照してください。
12. [表示] タブをクリックすると、リポジトリに入力されたさまざまなデータソースの詳細が表示され、既存のエンティティを取得する選択基準を設定します。詳細は、p.42 [データソースの表示 \(ストリーミング EA ノード\)](#) を参照してください。
13. ノードが正しく設定されたら、[OK] をクリックします。
14. テーブルノードをストリーミング EA ノードに接続してストリームを実行します。

テーブル ノードの出力ウィンドウには、データソースに新しいレコードを一致させる取得したすべてのエンティティを示しています。出力フィールドには、プレフィックス \$EA- が追加されます。詳細は、p.44 ストリーミング EA ノードの出力 を参照してください。

## 入力フィールドのフィーチャへのマッピング (ストリーミング EA ノード)

[入力] タブには、リポジトリのフィーチャにこのノードへの入力のフィールドををマッピングするためのオプションが含まれています。このタブまたは [表示] タブでマッピングの割り当てを設定し、リポジトリのすべてのデータ ソースの詳細を表示し、[OK] をクリックします。

一連のマッピングをすでに XML ファイルに保存している場合、[マッピングのエクスポート] をクリックして使用することができます。

図 3-11  
新規レコードのフィールドをリポジトリのフィーチャにマッピング



**エンティティ リポジトリ:** 存在する場合、現在のエンティティのリポジトリを示しています。複数存在する場合に別のリポジトリを選択するには、リストからいずれかを選択します。新しいリポジトリを作成するには、[<参照...>] を選択してダイアログ ボックスを表示し、リポジトリを作成します。詳細は、p.19 エンティティ リポジトリのオプション を参照してください。



**エンティティタイプ:** リポジトリに定義されるエンティティ タイプのリスト (機能のセット)。リストからエンティティ タイプを選ぶか、<新しいエンティティタイプを追加...> をクリックして [リポジトリ設定] ウィンドウを表示し、新しいエンティティ タイプを定義します。 詳細は、 p.25 [エンティティ リポジトリの構成](#) を参照してください。

**マッピング テーブル:** このテーブルでは、リポジトリ内の対応する機能への各入力フィールドをマッピングすることができます。適切な機能が選択したエンティティタイプに存在しない場合は、ここで新しい機能を作成することができます。

- **フィールド:** 選択したデータ ソース内の入力フィールドのセット。各フィールドには、フィールドの尺度 (データ型) を示すアイコンがあります。
- **機能にマップ:** フィールドを機能にマッピングするには、フィールドの行でこの列をダブルクリック (またはスペースバーを押す) し、リストから機能を選択します。適切な機能を使用できない場合、<新しいエンティティタイプを追加...> をクリックして [リポジトリ設定] ウィンドウを表示し、新しいエンティティ タイプを定義します。 詳細は、 p.25 [エンティティ リポジトリの構成](#) を参照してください。
- **使用方法:** 自宅と職場の電話番号など、複数のコンテキストが可能な特定のフィールドのコンテキストを示します。 詳細は、 p.31 [エンティティ タイプの管理](#) を参照してください。

**マッピングのインポート:** 以前エクスポートされたフィールドからフィーチャへのマッピングを外部 XML ファイルからインポートします。それは、異なるソースに対して同じマッピングを再定義する必要がなくなるため、同じマッピング要件を持つさまざまなデータソースがある場合に便利です。

**マッピングのエクスポート:** マッピング テーブルに表示されたフィールドからフィーチャへのマッピングのセットを外部 XML ファイルにエクスポートします。

## フィールド マッピングの表示 (ストリーミング EA ノード)

[リポジトリ] タブで、[リフレッシュ] をクリックし、入力フィールドがマッピングされたリポジトリ フィーチャを確認します。現在のデータソース (このエクスポートのノードに接続されたソースノードによって制御)、またはすべてのデータソースのいずれかのフィーチャを確認できます。

図 3-12  
フィールド マッピングの表示



**入力を表示：** 現在のデータ ソース、またはリポジトリに読み込まれたすべてのデータ ソースについてマッピングを表示するオプションを選択します。

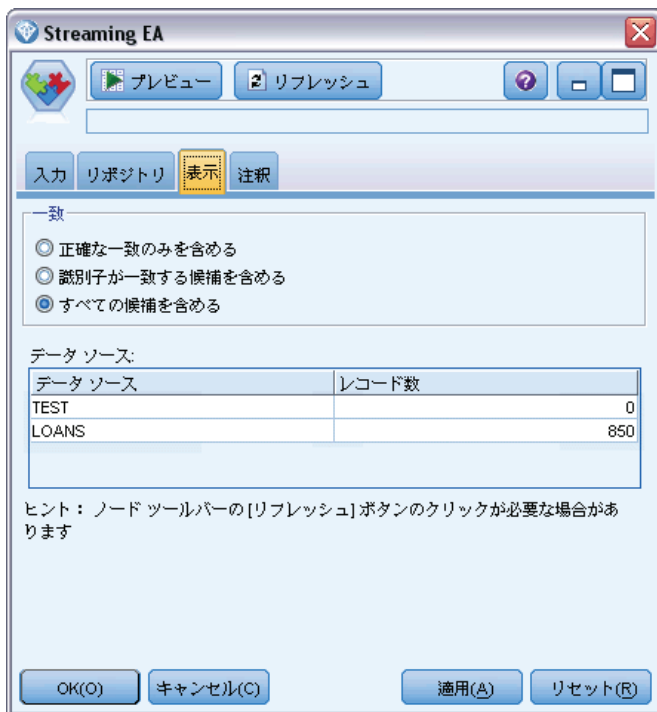
**フィーチャ。** 表示されたデータ ソースにマッピングがあるすべてのフィーチャのリスト。マッピングされていないフィーチャは表示されません。

**このノード：** 各列には、マッピングが定義されている各フィーチャの特定のデータ ソースのマップされたフィールドが表示されます。

## データ ソースの表示 (ストリーミング EA ノード)

[表示] タブでは、リポジトリに入力されたさまざまなデータソースの詳細を見ることができます。このノードへの入力が処理されるデータ ソースで、一致するエンティティを検索および取得します。レコード数を更新するには、[リフレッシュ] をクリックします。

図 3-13  
データソースの詳細の表示



**一致:** これらのオプションは、[入力] タブで指定したフィールド ツーフィーチャのマッピング情報が候補レコード（つまりリポジトリ全体の内容）と照合する方法を厳密に指定します。一致基準に近いほど、取得するエンティティは少なくなります。

- **完全一致のみ含める:** これは、最も近い一致基準で、選択されるレコードが最も少なくなります。完全一致と見なされるエンティティのみを返す場合、このオプションを使用します。
- **一致する識別子の候補を含める:** 同じ ID（クレジットカード番号や、納税番号など、度数 1 で設定したフィーチャの ID）を共有するエンティティを返す場合にこの設定を使用します。
- **すべての候補を含める:** フィーチャを共有している最大限のエンティティを表示する場合にこのオプションを使用します。これは、最もゆるい一致基準で、選択されるレコードが最も多くなります。このオプションは、完全一致またはほぼすべてのフィーチャを共有するエンティティ（通常 1 または少数の度数で設定）を返します。たとえば、同じ納税番号のエンティティや、住所が似ているエンティティなどです。

**データソース:** この表には、各ソース内のレコード数とともに、リポジトリに読み込まれたさまざまなデータソースのソースタグが示されています。

## ストリーミング EA ノードの出力

ストリーミング EA ノードの出力は、取得された各レコードの以下のフィールドで構成されています。

フィールド	説明
Field1[, Field2[, ...FieldN]]	新しいレコードを含むデータ ソースのフィールド。
\$EA-ID	リポジトリのこのレコードのエンティティ識別子。
\$EA-SC	このレコードとリポジトリで観測されたエンティティ間の一致のフィールドを示す親密さ。1.0（一致が少ない）～10.0（一致が多い）の値。
\$EA-SRC	レコードが作成元であるデータ ソースを示すソース タグ
\$EA-KEY	データソースファイル内のこのレコードの一意キーの値。
\$EA-Feature1[, \$EA-Feature2[, ...\$EA-FeatureN]]	リポジトリのこのマップされているフィーチャの値。

## 他のIBM SPSS製品との IBM SPSS Modeler Entity Analytics の使用

インストーラは、次の製品と IBM SPSS Modeler Entity Analytics を使用することができます。

- IBM SPSS Collaboration and Deployment Services
- IBM SPSS Modeler Batch for Windows
- IBM SPSS Modeler Solution Publisher

これらの製品と IBM SPSS Modeler Entity Analytics の機能を使用する前に、これらのインストーラを実行する必要があります。詳細については、『IBM SPSS Modeler Premium インストール ガイド』を参照してください。

## 管理タスク

Entity Analytics 内に作成されたリポジトリの場合、新しいデータベース・サービスは、IBM solidDB 製品を使用して作成されます。solidDB に関連する管理タスクがあります。これらのタスクは通常、データベース管理者またはシステム管理者によって実行されます。

- ポート割り当ての設定
- リポジトリ データベースの管理者の資格情報の管理

実行する必要がある管理タスクはすべてのリポジトリに手供されます。タスクは次のとおりです。

- リポジトリ ストレージ ディレクトリの移動
- 日付/時刻およびタイムスタンプのフィールドのストリーム オプションの設定
- 同じ Windows システムでの SPSS Modeler クライアントと SPSS Modeler Server との IBM SPSS Modeler Entity Analytics の実行
- エンティティ リポジトリの消去
- エンティティ リポジトリの削除
- リポジトリに接続できない場合にリポジトリを削除する

## ポート割り当ての設定

各 solidDB のデータベース サービスは、マシン上で実行されている他のサービスに割り当てることができないポートを割り当てる必要があります。データベース サービスは、IBM® SPSS® Modeler Server を実行する同一マシン（または、IBM® SPSS® Modeler が SPSS Modeler Server に接続せずに使用されている場合は、IBM® SPSS® Modeler を実行するマシン）上にあります。

デフォルトでは、Entity Analytics は 1320 ~ 1520 のポートを割り当て、作成される最初のリポジトリに 1320 を割り当てます。競合する場合、ファイル <modeler server installation path>/ext/bin/pasw.entityanalytics/ea.cfg を編集し、min\_port 設定および max\_port 設定に適切な値を設定して、ポートの割り当てを設定することができます。このファイルのデフォルトのコンテンツは次のとおりです。

```
# port range configuration for entity analytics
#
# this port range controls which ports SolidDB databases
# (created to store Entity Analytics Repositories in)
# may use. Configure this if the default port range will
# introduce a conflict on your system.
#
# default min_port = 1320
# default max_port = 1520
```

min\_port, 1320

max\_port, 1520

## リポジトリ データベースの管理者の資格情報の管理

リポジトリが作成されると、エンティティ リポジトリをホストしている solidDB データベースの管理者ユーザー名およびパスワードが定義されます。現在の資格情報が分かっている場合、solidDB SQL エディタを使用してこれらの詳細情報を変更できます。

### solidDB SQL エディタを起動するには

1. クライアント マシンで、コマンド プロンプト ウィンドウを開きます。
2. 以下を入力します。

```
cdmodeler_install_dir\ext\bin\pasw.entityanalytics\solidDB\bin
```

modeler\_install\_dir は、SPSS Modeler がインストールされているディレクトリです。

3. 以下を入力します。

```
solsql -c "C:\Documents and Settings\All Users\Application  
Data\IBM\SPSS\Modeler\version\EA\repositories\repos_name
```

version は、SPSS Modeler のインストールのバージョン番号で、repos\_name はリポジトリの名前です。

4. プロンプトの後、現在のデータベース管理者のユーザー名およびパスワードを入力し、solsql> プロンプトを表示します。

### データベース管理者のパスワードを変更するには

1. solsql> プロンプトで、次のコマンドを入力します。

```
alter userusername identified bypassword;  
commit work;
```

username はデータベース管理者の現在のユーザー名、password は新しいパスワードです。

2. exit; と入力してエディタを閉じます。
3. SPSS Modeler クライアントを再起動します。

### その他の管理タスクを実行するには

solidDB データを扱うその他の管理タスクの詳細については、該当するバージョンの IBM solidDB のマニュアル (<http://publib.boulder.ibm.com/>) を参照してください。

## リポジトリ ストレージ ディレクトリの移動

デフォルトでは、リポジトリ ファイルは以下の場所の EA という名前のディレクトリに保存されます。

- C:\Documents and Settings\All Users\Application Data\IBM\SPSS\Modeler\version\EA (Windows systems)
- modeler\_install\_directory/ext/bin/pasw.entityanalytics/EA (UNIX systems)

リポジトリの保存に使用されるファイルは非常に大きくなるため、異なるディスク又はパーティションに移動して、より多くの領域を使用できるようにする必要があります。そのためには、以下の手順に従います。

1. SPSS Modeler を終了します。
2. EA ディレクトリを元の場所（上記参照）から新しい場所に移動します。たとえば、Windows では F:\data\EA などの新しい場所に移動します。
3. ファイル <modeler server installation path>/ext/bin/pasw.entityanalytics/ea.cfg を編集して以下のオプションを追加します。

```
repository_data_directory,new_location
```

new\_location は EA ディレクトリの移動先となるディレクトリ（例：F:\data\EA）です。

## 日付/時刻およびタイムスタンプのフィールドのストリーム オプションの設定

ソース データに日付/時刻またはタイムスタンプ データを含むフィールドが含まれる場合、対応するストリーム オプションが IBM SPSS Modeler Entity Analytics によって認識された形式に設定されます。

### ストリーム オプション形式を設定するには

1. SPSS Modeler のメイン メニューで次の各項目を選択します。  
ツール > ストリームのプロパティ > オプション :
2. [日付/時刻] を選択します。

3. [日付の形式] を [YYYY-MM-DD] に設定します。
4. [時刻の形式] を [HH:MM:SS] に設定します。
5. [OK] をクリックします。

## 同じ Windows システムでの SPSS Modeler クライアントと SPSS Modeler Server との IBM SPSS Modeler Entity Analytics の実行

IBM SPSS Modeler Entity Analytics を同一 Windows システムの SPSS Modeler クライアント SPSS Modeler Server にインストールしている場合、デフォルトではクライアントとサーバーは同じリポジトリで共有します。別のリポジトリを使用する場合、システムのうち **1 つ** の構成ファイル `ea.cfg` を編集して、異なるポート番号及びリポジトリ フォルダを使用する必要があります。

注：特に、32 ビットの SPSS Modeler クライアントおよび 64 ビット SPSS Modeler Server（又はその逆）を使用する場合、この手順を実行する必要があります。

1. ファイル `<modeler [server] installation path>/ext/bin/pasw.entityanalytics/ea.cfg` 開いて編集します。
2. 他のシステムの別のポートを使用するよう、`min_port` および `max_port` 設定を変更します。詳細は、p. 45 [ポート割り当ての設定](#) を参照してください。
3. 他のシステムの別のディレクトリを使用するよう、`repository_data_directory` 設定を変更します。
4. `ea.cfg` ファイルを保存して閉じます。

## エンティティ リポジトリの消去

エンティティのリポジトリからデータ レコードを消去しつつ、設定情報を維持したい場合、リポジトリの消去オプションを使用します。

### リポジトリを消去するには

1. エンティティ分析ノードを開きます。
2. [エンティティリポジトリ] リストをクリックします。
3. <[参照...]> をクリックして、[エンティティ解決インスタンス] ダイアログ ボックスを開きます。



4. [エンティティ解決インスタンス] ダイアログボックスで、[リポジトリ名] リストをクリックします。
5. 消去するリポジトリを選択します。
6. 接続されていない場合、管理者のユーザー名とパスワードを入力し、[接続] をクリックします。
7. [リポジトリの消去] ボタンが有効な場合、クリックします。
8. [データ ソースの消去] ダイアログの [リポジトリの消去] をクリックして、リポジトリの消去を確定します。

## エンティティ リポジトリの削除

リポジトリが必要でない場合、完全に削除することができます。

注意: 上記のとおり、この操作でリポジトリは完全に削除されます。この操作は取り消しできません。不安な場合は、[消去] ボタンを使用してすべてのソース データを削除します。データは削除されても、リポジトリの設定は削除されません。詳細は、 p. 48 [エンティティ リポジトリの消去](#) を参照してください。

注： 次の手順では、SPSS Modeler からリポジトリに接続でき、リポジトリをホストするデータベースの管理者ユーザー名とパスワードを知っていることを前提としています。そうでない場合、リポジトリに接続できないときはリポジトリを削除する手順を行います。詳細は、 p. 50 [リポジトリに接続できない場合にリポジトリを削除する](#) を参照してください。

### リポジトリを削除するには

1. エンティティ分析ノードを開きます。
2. [エンティティリポジトリ] リストをクリックします。
3. <[参照...]> をクリックして、[エンティティ解決インスタンス] ダイアログ ボックスを開きます。
4. [エンティティ解決インスタンス] ダイアログボックスで、[リポジトリ名] リストをクリックします。
5. 削除するリポジトリを選択します。
6. 接続されていない場合、管理者のユーザー名とパスワードを入力し、[接続] をクリックします。
7. [リポジトリ全体の削除] ボタンが有効な場合、クリックします。
8. [削除] をクリックしてリポジトリの削除を確定します。

9. [OK] をクリックして、削除が正常に行われたことを確認します。

## リポジトリに接続できない場合にリポジトリを削除する

SPSS Modeler からの接続の問題により、またはユーザー名かパスワードを忘れたためにエンティティ リポジトリを削除したいが接続できない場合、次の手順を行います。

リポジトリ データベースをホストするマシンでこの手順を実行します。

### Windows システム

1. コマンド プロンプト ウィンドウを開きます。
2. 以下を入力します。

```
cdmodeler_install_dir  
cd ext\bin\pasw.entityanalytics\tools  
delete_repository.batrepos_name
```

modeler\_install\_dir は、SPSS Modeler がインストールされているディレクトリ、repos\_name はリポジトリの名前です。

3. このセクションの「手順の完了」から続行します。

### UNIX システム

1. シェルを開きます。
2. 以下を入力します。

```
cdmodeler_server_install_dir  
cd ext/bin/pasw.entityanalytics/tools  
./delete_repository.shrepos_name
```

modeler\_server\_install\_dir は、SPSS Modeler Server がインストールされているディレクトリ、repos\_name はリポジトリの名前です。

### 手順の完了 (すべてのシステム)

1. プロンプトで、Y を入力してリポジトリの削除を確定します。
2. リポジトリが削除されると、以下のようなメッセージが表示されます。

```
Info - please remove repository files from the following directory:  
directory_path  
(Note - you may need to reboot first before you can remove the repository files)
```

3. 削除したリポジトリと同じ名前のディレクトリを削除します。ディレクトリを削除できない場合、マシンを再起動してもう一度実行してください。

# 実際のエンティティ分析

## この例について

この例では、IBM® SPSS® Modeler を使用して収録できる印象的な結果に対してエンティティ分析の追加をどのように改善できるかについて説明します。

この例では、データ ファイル `loan_applications.csv` を参照するストリーム `loan_entity_analytics.str` を使用します。これらのデータ ファイルは、IBM® SPSS® Modeler Entity Analytics もインストールしている SPSS Modeler のインストール ディレクトリ中の `Demos` ディレクトリにあります。Windows の [スタート] メニューの SPSS Modeler プログラム グループから `Demos` ディレクトリにアクセスできます。`loan_entity_analytics.str` ファイルは `Entity_Analytics` ディレクトリにあります。

注： この例ではストリームを実行する前に、システム上にリポジトリを作成する必要があります。この例を続行する前に行ってください。詳細は、3 章 p.17 [リポジトリの作成](#) を参照してください。

それでは、銀行は、申し込みが保留されているローンに対して顧客が滞納する可能性があるかどうかを懸念しているというシチュエーションについて考えてみましょう。銀行の IT 部門は、SPSS Modeler を長い間使用しているため、すでにストリームを作成し、銀行が過去に行った融資についての既存のデータから予測モデルを構築しています。これらのローンについては返済されているか、顧客は返済を怠っています。

## 元のモデル

ここで、どのようにモデルを構築し、そこから何が分かるかについて説明します。

図 4-1  
モデリング ノードの初期ストリーム



過去のローンの詳細と同様、loan\_applications.csv データ セットには、ローンの申し込みが保留中である 150 名の顧客の詳細、合計 850 件のレコードが含まれています。

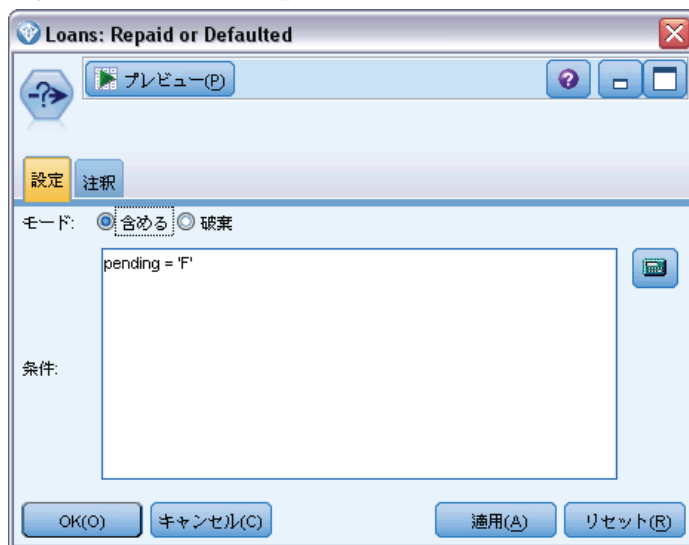
データ セットのフィールドのすべてが、予測を行う際に有用であるではありません。たとえば名前のフィールドは無視できます。データ型ノードは、役割を[なし]に設定して無視するフィールドを除外します。予測を行うために使用するフィールドは、役割を[入力]に設定し、予測を行うモデルの値のフィールドは[対象]に設定します。

図 4-2  
データ型ノードに設定するフィールドの役割



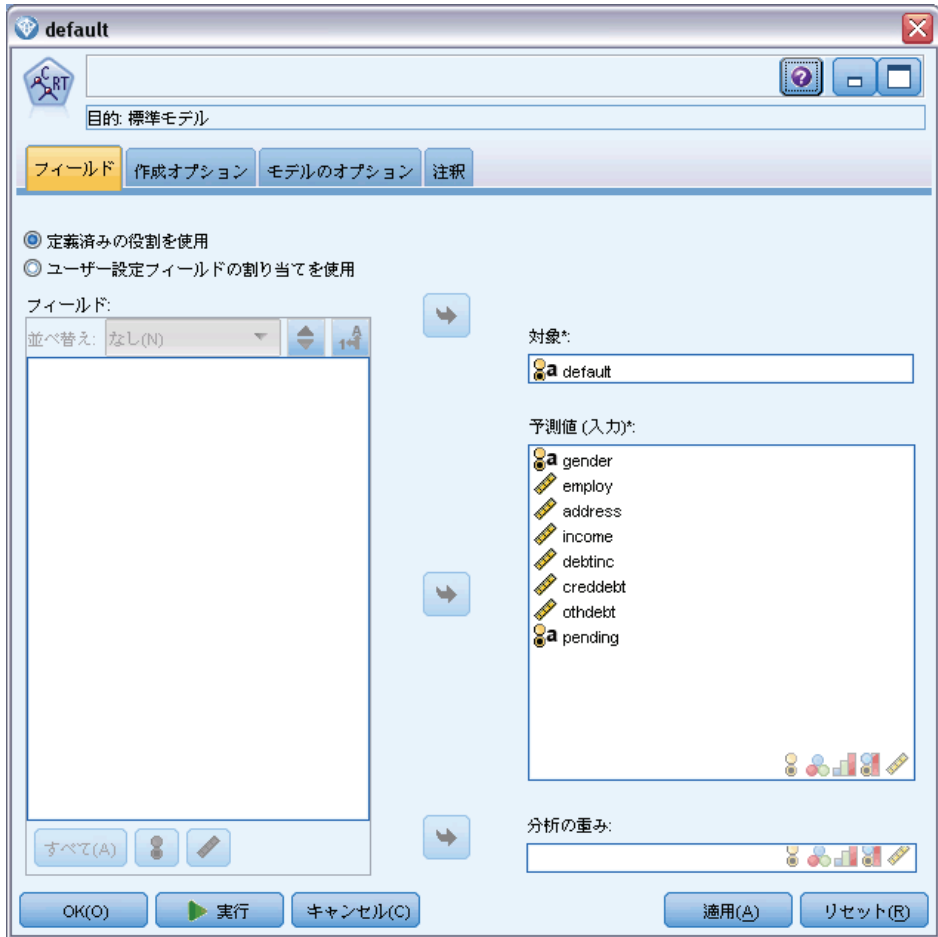
モデルは過去のデータのみに基づいて予測を行うため、ストリームには保留中としてマークされていないローンだけを含ま条件抽出ノードが含まれるため、150 件の保留中のローンが破棄されます。

図 4-3  
保留中のローン申し込みを破棄



破棄されている保留中のローンで、返済済みまたは滞納しているローン 700 件については、その詳細がモデリング ノードに渡されます。銀行はさまざまな SPSS Modeler アルゴリズムのいずれかを使用して、最適なモデルを作成することができます。この場合、C&R Tree ノードを使用して、銀行の顧客の過去の返済状況に基づいて滞納の可能性を予測するモデルを構築しています。

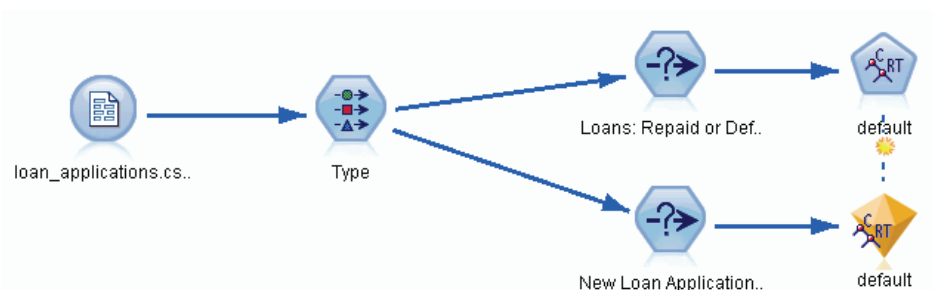
図 4-4  
予測値と対象フィールドの割り当て



予測に使用されるフィールドは予測フィールドとして、モデルがこの場合は滞納を予測しようとする値のフィールドは、以前にデータ型ノードで定義したように、対象フィールドとして設定されます。

このストリームを実行すると、予測フィールドから構築されたモデルを含む、モデル ナゲットを生成します。

図 4-5  
モデル ナゲットを追加したストリーム



銀行のアナリストは、モデルを使用して保留中の返済を持つ顧客が滞納する可能性があるかどうかについての予測を開始することができます。元のデータセットを使用して、アナリストは、破棄するのではなく、保留中としてマークされた 150 件のローンのレコードのみを含む条件抽出ノードを挿入します。アナリストは、モデルの予測を視覚的に表現するための棒グラフ ノードを追加して、直接モデルにこれらのレコードを渡します。

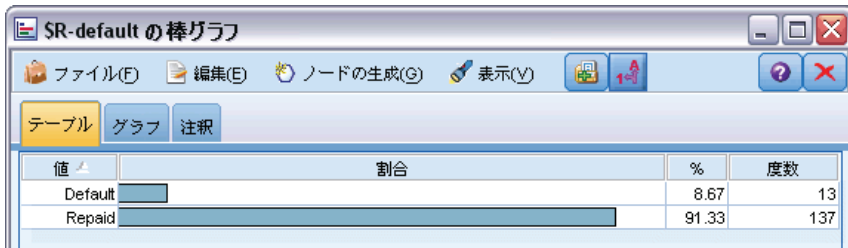
図 4-6  
新しいローンの申し込みを選択し、棒グラフノードを追加したストリーム



棒グラフノードは、モデルの  $\$R$ -default フィールドの値の分布を示します。このフィールドは、実行時に C&R Tree ノードによってデータモデルに追加されます。フィールドには、新しい申請者が返済するか債務不履行があるかに関する予測が含まれ、このフィールドを後で使用してエンティティ分析の追加の影響を比較します。

ストリームのこの部分を実行し、アナリストは 150 件の新規申し込みのうちの 137 件が、そのローンの返済が予測されるということを棒グラフノードの出力から分かります。残りの 13 件は滞納が予測されるため、アナリストは銀行にこの申し込みを却下するよう推奨します。

図 4-7  
エンティティ分析のない棒グラフノードの出力



## エンティティ分析の追加

等式にエンティティ分析を加えることによって状況が改善することができるかどうか、見てみましょう。あなたは、ソース データの顧客レコードで不正の構成があるエントリを調査する、エンティティ分析の専門家であると想像してください。データ入力エラーが原因で重複レコードがある可能性はありますが、ローンの申込者が自分の身元を偽装しようとする可能性もあります。いずれにせよ、銀行は実態を知る必要があります。

この例の目的のため、エンティティのリポジトリがすでに作成されていることを前提としています。詳細は、3 章 p.17 [リポジトリの作成](#) を参照してください。

## ソース データをリポジトリに取得する

まず、EA エクスポート ノードをデータ ソース ノードに追加して、ソース データをエンティティ リポジトリに追加できるようにします。

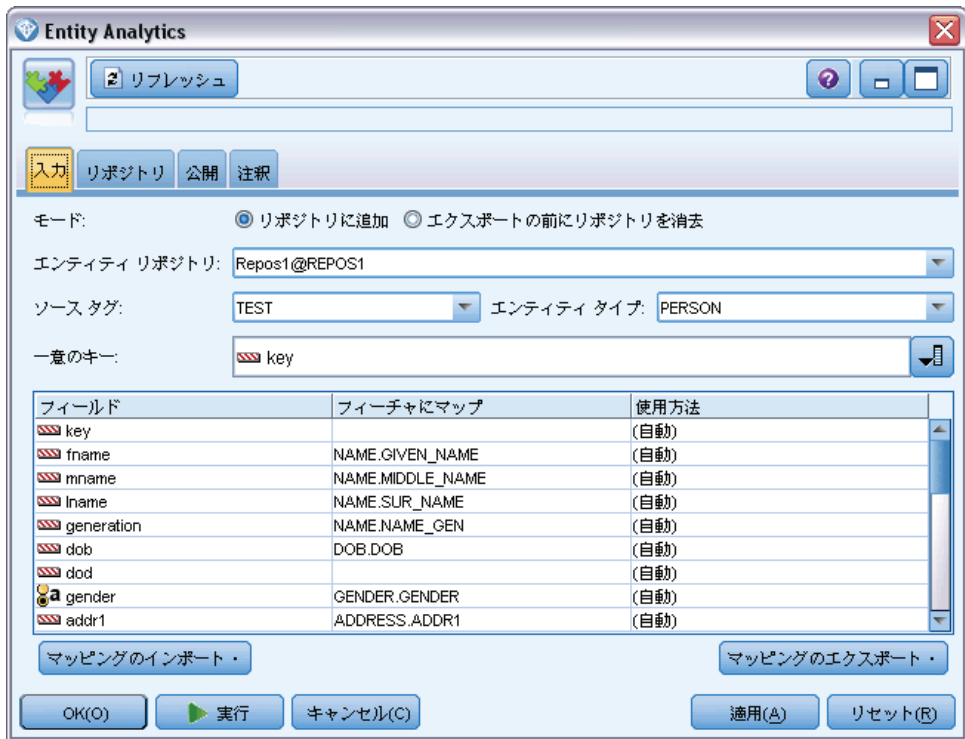
図 4-8  
EA 分析エクスポート ノードのデータソース ノードへの接続



データをエクスポートする前に、データ ソースのフィールドをエンティティ リポジトリのフィーチャにマッピングする必要があります。これは、さまざまなデータ ソースが同じ種類の情報に対して異なるフィールド名を使用する場合があるためです。エンティティ リポジトリは、重複を避けるために、情報タイプの標準セット（「フィーチャ」）を提供します。



図 4-9  
フィールドのフィーチャへのマッピング



EA エクスポート ノードで、リポジトリに関する詳細を設定します。接続の詳細、ソース タグ（この場合データ ソース - TEST を識別）、エンティティ タイプ（PERSON という名前のフィーチャのセット）、一意のキーフィールド（各レコードを一意に識別）を設定します。この場合、キーフィールドを一意キーとして使用します。

マッピングを設定することができます。使用しているフィーチャ セットに、フィールド 姓、名、年代、生年月日、性別、住所1、市区町村、国、郵便番号、電話番号、電子メール、社会保障番号、運転免許証番号、パスポート番号に対応するフィーチャがあります。

まず、姓のマッピングを設定します。[姓] 行のテーブルの [フィーチャにマップ] 列をダブルクリックして NAME.GIVEN\_NAME エントリまで下方向にスクロールし、クリックしてマッピングを作成します。

対応するフィーチャを持つ残りのフィールドをマッピングし、マッピングの設定がこのようになるようにします。

テーブル 4-1  
リポジトリのフィーチャにマッピングしたフィールド

フィールド	フィーチャにマップ
fname	NAME. GIVEN_NAME
mname	NAME. MIDDLE_NAME
lname	NAME. SUR_NAME
generation	NAME. NAME_GEN
生年月日	DOB. DOB
性別	GENDER. GENDER
addr1	ADDRESS. ADDR1
city	ADDRESS. CITY
country	ADDRESS. COUNTRY
postcode	ADDRESS. POSTAL_CODE
電話番号	PHONE. PHONE_NUM
電子メール	EMAIL_ADDR. ADDR
社会保障番号	SSN. ID_NUM
運転免許証番号	DRLIC. ID_NUM
パスポート番号	PASSPORT. ID_NUM

[実行] をクリックしてデータをリポジトリにエクスポートします。この作業時は時間がかかります。[実行のフィードバック] ダイアログ ボックスが閉じると、エクスポートが完了します。

## 解決した識別の読み込み

データをリポジトリにエクスポートすると、エンティティ分析システムは \$EA-ID フィールドとして表示される一意のエンティティ識別子を割り当て、識別の競合の可能性を解決します(注：これは、フィールドがデータソース レコードの一意の識別に使用される EA エクスポート ノードの一意のキー フィールドと同じではありません)。

解決した識別を読み込む最初のステップは、エンティティ (EA) 入力ノードをストリームに追加することです。この入力ノードは、この段階でどこにも接続する必要はありません。

図 4-10  
エンティティ分析 (EA) 入力ノード



Entity Analytics(EA)

エンティティ分析 (EA) 入力ノードで、エンティティ リポジトリの詳細を設定します。リポジトリにエクスポートされたデータ ソースのリストが表示されます。この場合、データ ソースは 1 つだけです。

図 4-11  
データ ソースをリポジトリで選択



[TEST] データ ソース のチェックボックスをオンにして、[OK] をクリックします。

エンティティ分析システムがデータにどのような処理をしたのかを見てみましょう。テーブル ノードをエンティティ分析 (EA) 入力ノードに接続し、テーブル ノードを開いて [実行] をクリックし、テーブル ノード出力ウィンドウを開きます。

図 4-12  
テーブル ノードの出力

	\$EA-ID	\$EA-SRC	\$EA-KEY
1	1	TEST	REF00000001
2	2	TEST	REF00000002
3	3	TEST	REF00000003
4	4	TEST	REF00000004
5	5	TEST	REF00000005
6	6	TEST	REF00000006
7	7	TEST	REF00000007
8	8	TEST	REF00000008
9	9	TEST	REF00000009
10	10	TEST	REF00000010
11	11	TEST	REF00000011
12	12	TEST	REF00000012
13	13	TEST	REF00000013
14	14	TEST	REF00000014
15	15	TEST	REF00000015
16	16	TEST	REF00000016
17	17	TEST	REF00000017
18	18	TEST	REF00000018
19	19	TEST	REF00000019
20	20	TEST	REF00000020

1 つのフィールドに \$EA-KEY のラベルが付いています。これは実際ソースデータの キー フィールドです。これは EA エクスポート ノードで [一意のキー] フィールドとして選択したためです。

システムは他に 2 つのフィールドを追加しています。\$EA-ID フィールドはソース レコードではなく、解決された識別子の一意の識別子です。すぐに異なって表示されます。\$EA-SRC フィールドは、データがどこから由来するものかを示します。ここでは、EA ノードで割り当てたソース タグであるため [TEST] と表示します。

ソース データの他のフィールドについてはどうでしょう?それらはリポジトリ内にありますが、パフォーマンス上の理由により、エンティティ分析 (EA) 入力ノードは、最小限のフィールドを下流に渡して処理します。

テーブル ノード出力を 385 行目までスクロールします。

図 4-13  
テーブル出力の行と \$EA-ID 番号の相違点

	\$EA-ID	\$EA-SRC	\$EA-KEY
376	376	TEST	REF00000376
377	377	TEST	REF00000377
378	378	TEST	REF00000378
379	379	TEST	REF00000379
380	380	TEST	REF00000380
381	381	TEST	REF00000381
382	382	TEST	REF00000382
383	383	TEST	REF00000383
384	384	TEST	REF00000384
385	45	TEST	REF00000385
386	386	TEST	REF00000386
387	387	TEST	REF00000387
388	388	TEST	REF00000388
389	389	TEST	REF00000389
390	390	TEST	REF00000390
391	391	TEST	REF00000861
392	392	TEST	REF00000392
393	393	TEST	REF00000393
394	394	TEST	REF00000394
395	395	TEST	REF00000395

ここで、\$EA-ID 番号が順番を外れてどのように表示されるのかが分かります。エンティティ システム分析では、レコード REF00000385 が、レコード REF0000045も持ち、エンティティ 45 として識別される個人を参照しすることを確認しました。出力をさらにスクロールすると、485 行目、517 行目、520 行目など、より多くの番号が順番を外れています。さらに詳細に確認する必要があります。

まず、キーに名前を変更して、\$EA-KEY フィールドにソース データのキーフィールドのデータが含まれていることを強調します。フィルタ ノードをエンティティ分析 (EA) 入力ノードに接続し、フィルタ ノードを開きます。2 番目の [フィールド] 列のワード \$EA-KEY をダブルクリックし、key と入力します。

図 4-14  
\$EA-KEY フィールドの名前変更



[OK] をクリックして、フィルタ ノードを閉じます。

\$EA-ID エンティティ ID を昇順にソートします。ソート ノードをフィルタ ノードに接続します。ソート ノードを開き、[ソート基準] テーブルの隣にある最上位のボタンをクリックし、\$EA-ID を選択して [OK] をクリックします。

図 4-15  
エンティティ ID を昇順にソート



ソート順を [昇順] のままにし、[OK] をクリックします。

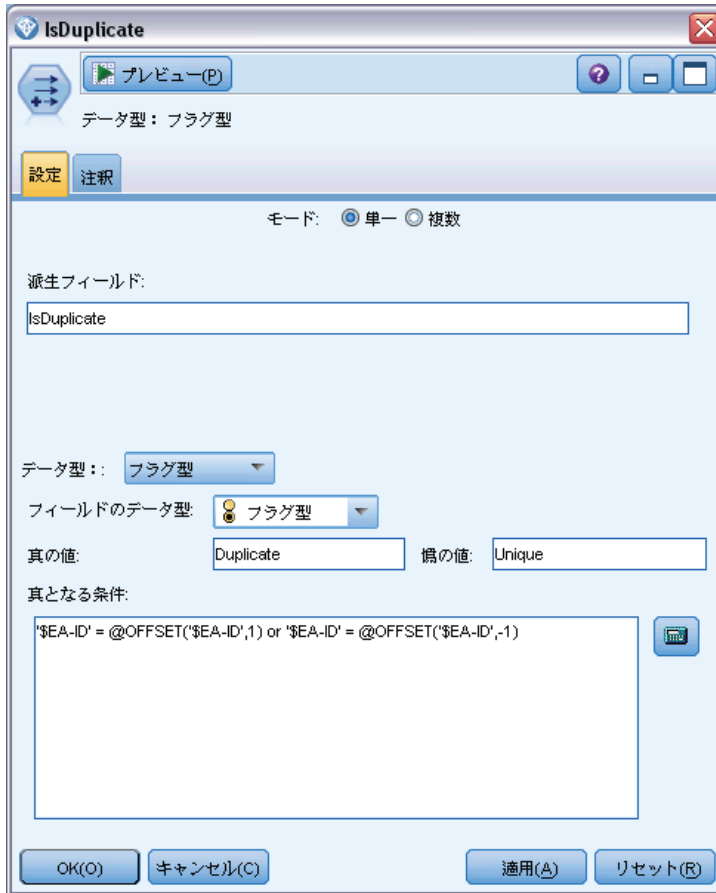
レコードが一意か重複しているかを示す追加のフィールドを作成する必要があります。フィールド作成ノードをこのソート ノードに接続します。フィールド作成ノードを開き、[派生フィールド] 名を `IsDuplicate` に設定します。[データ型] リストから、[フラグ型] を選択し、[フィールドタイプ] を [フラグ型] に設定します。[True の値] フィールドを [重複] に設定し、[False の値] フィールドを [-一意] に設定します。

重複したレコードを探すには、SPSS Modeler で提供されている `@OFFSET` という名前の特別なシーケンス関数を利用します。

[If] フィールドに以下を入力します。

```
'$EA-ID' = @OFFSET('$EA-ID',1) or '$EA-ID' = @OFFSET('$EA-ID',-1))
```

図 4-16  
フィールド作成ノードでの条件設定



エンティティ ID を昇順にソートした状態で、@OFFSET 関数は 2 つの隣接するエンティティ ID が同一であるかどうか、その場合は 2 つのレコードが重複していることを確認します。その場合、IsDuplicate 値が Duplicate に設定され、そうでない場合は Unique に設定されます。

[OK] をクリックして、ノードを閉じます。

フィールド作成ノードの影響を確認するには、テーブル ノードをフィールド作成ノードに接続し、テーブル ノードを開いて [実行] をクリックします。テーブル ノード出力ウィンドウを 45 行目までスクロールします。



図 4-17  
フィールド作成ノードの出力

	\$EA-ID	\$EA-SRC	key	IsDuplicate
39	39	TEST	REF00000039	Unique
40	40	TEST	REF00000040	Unique
41	41	TEST	REF00000041	Unique
42	42	TEST	REF00000042	Unique
43	43	TEST	REF00000043	Unique
44	44	TEST	REF00000044	Unique
45	45	TEST	REF00000045	Duplicate
46	45	TEST	REF00000385	Duplicate
47	46	TEST	REF00000046	Unique
48	47	TEST	REF00000047	Unique
49	48	TEST	REF00000048	Unique
50	49	TEST	REF00000049	Unique
51	50	TEST	REF00000050	Unique
52	51	TEST	REF00000051	Unique
53	52	TEST	REF00000052	Unique
54	53	TEST	REF00000053	Unique
55	54	TEST	REF00000054	Unique
56	55	TEST	REF00000055	Unique
57	56	TEST	REF00000056	Unique
58	57	TEST	REF00000057	Unique

エンティティ分析 (EA) ノードからの直接の出力を思い出してください。システムはすでに REF00000385 がエンティティ 45 と同じ人物を示していることを認識しています。ここでは、REF00000045 と REF00000385 が重複して同じエンティティ 45 を示している事実にご注意します。

出力ウィンドウをさらに下にスクロールし、重複としてフラグが付いている他のレコードを確認します。

重複したレコードを一覧表示するレポートを取得するには、レポート ノード (ノード パレットの [出力] タブ) を [IsDuplicate] フィールド作成 ノードに接続します。レポート ノードを開き、[テンプレート] タブの入力フィールドに以下のテキストをコピーし、[実行] をクリックします。

```
<html>
<h1>重複した顧客レコードのリスト

<h2>次のレポートが生成されました。[@TODAY]

<h2>Duplicate records
<table>
<tr>
<td>Entity ID</td>
```

```
<td>Key</td>
</tr>
```

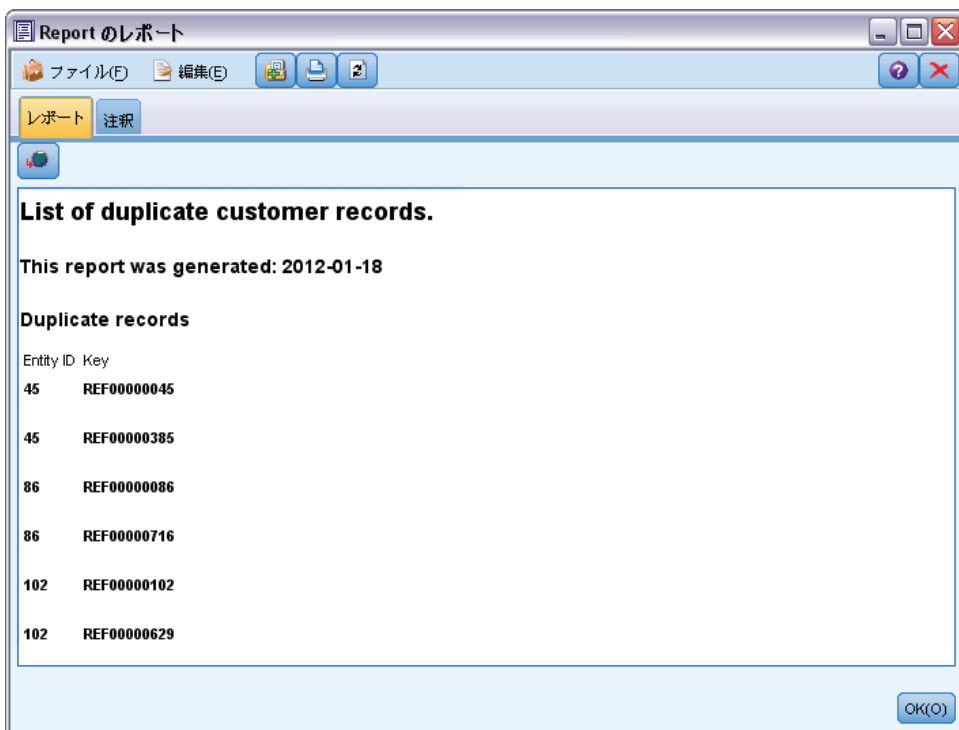
```
#WHERE IsDuplicate = "Duplicate"
```

```
<tr>
  <td>['$EA-ID']</td>
  <td>[key]</td>
</tr>
#
</table>
```

```
</html>
```

これにより、以下の出力が作成されます。

図 4-18  
レポート ノードの出力



この場合、HTML 形式を使用しますが、XML または ASCII 形式も使用することができます。

## エンティティ分析出力と元のモデルとの比較

この例の最終段階では、エンティティ分析を追加することにより、銀行が始めに行った予測に違いをもたらすかどうかを確認します元のモデルでは、150 件の保留中の申請のうち 13 名が滞納者であると予測しました。結合ノードを使用して、モデルの出力とエンティティ分析の重複レコードに関する情報を結合し、予測に変化をもたらすかどうかを確認します。

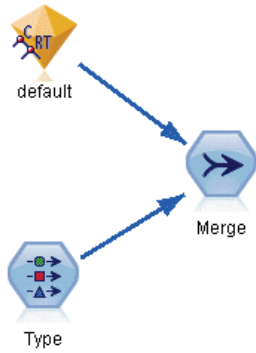
まず、エンティティ分析によって追加される新しいフィールドのデータ型が正しいこと、またはSPSS Modeler で認識されているような尺度であることを確認する必要があります。データ型ノードを IsDuplicate フィールド作成ノードに接続し、データ型ノードを開いて [値を読み込む] ボタンをクリックします。

図 4-19  
データ型ノードの設定



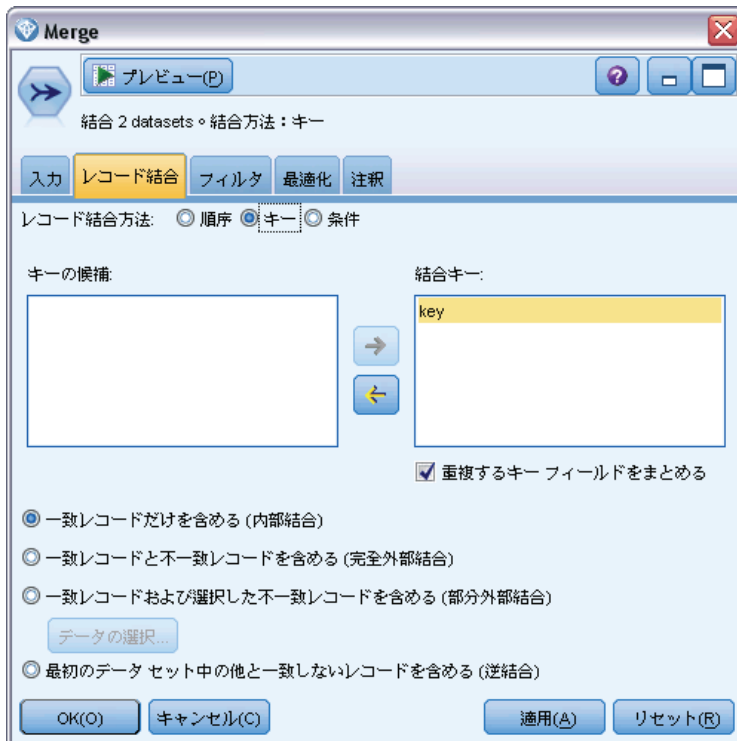
結合ノードを追加できます。データ型ノードに接続し、元のモデルを含む金色のナゲットにも接続します。金のナゲットに接続するには、[接続] を選択し、2 つの入力矢印が表示された結合ノードをクリックします。

図 4-20  
結合ノードへの入力



結合ノードを開き、[結合方法] を [キー] に設定して、右方向矢印ボタンをクリックし、キー フィールドを [可能なキー] から [結合するキー] に移動し、[OK] をクリックします。

図 4-21  
結合操作するキー フィールドを指定



これで、比較する準備がほぼ整いました。ただし、棒グラフノードを接続してこの時点で実行した場合、元の予測からの変更を確認できません。ストリームは、新しいフィールドがエンティティ分析によって作成された元のモデル ナゲットの出力を結合しますが、データ モデルの予測フィールド自体（\$R-default）は新しい情報で更新されていません。

このためには、置換ノードを使用してフィールド値を置き換えることができます。置換ノードを結合ノードに接続して、置換ノードを開きます。

[フィールドに入力] の右側の最上部のボタンをクリックし、リストの最下部までスクロールして [ \$R-default ] を選択、[OK] をクリックします。このフィールドは、ダイアログ ボックスの残りの部分で指定された条件が入力されている場合値が変更されます。

条件を指定するには、[置換] が [条件に基づく] に設定されていることを確認し、[条件] フィールドに以下を入力します。

```
default != "default" and IsDuplicate = "Duplicate"
```

[置換文字列] フィールドに以下を入力します。

```
"default"
```

図 4-22  
フィールド値を置き換える条件を指定



これらの設定には少し説明が必要です。条件では、元のデータ セットのデフォルト フィールドの値がデフォルトではなく重複しているものとしてフラグが付いているレコードについては、モデルの [\$R-default] フィールドの値はデフォルトに設定されるとしています。

[\$R-default] フィールドは、顧客がローンを滞納する可能性が高いかどうかについての予測を含むモデルのフィールドです。このようにして、レコードが重複した顧客は、潜在滞納者としてモデルに追加されます。

[OK] をクリックして、置換ノードを閉じます。

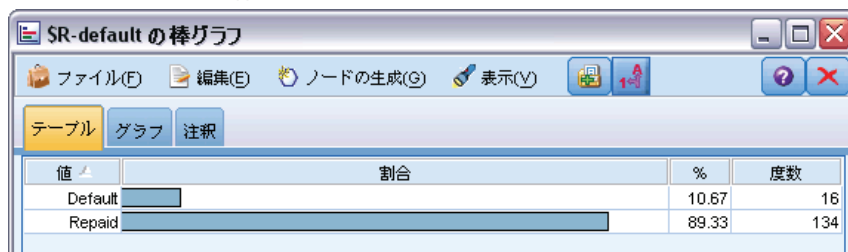
エンティティ分析による相違点を確認できるようになりました。[グラフ] パレットから、棒グラフ ノードを置換ノードに接続し、棒グラフ ノードを開きます。[フィールド] リストをクリックして、[\$R-default] を選択します。

図 4-23  
棒グラフ ノードの設定



[実行] をクリックして新しい予測のグラフを作成します。

図 4-24  
エンティティ分析後の棒グラフノードの出力



13 件の申請を除く 16 件がリスクのある申請です。彼らが滞納した場合、超過したこの 1 件が犠牲となるため、エンティティ分析をリスク評価業務に加えることの利点を実証することができます。

## 要約

この例では、エンティティ分析を使用して人物や組織に関するデータの重複レコードを除外し、予測品質の改善を行う方法について説明しました。

注：他の処理を行う前に重複レコードを除去することをお勧めします。自動データ準備（ADP）ノードを使用して、データ分析、固定値の識別、問題のあるまたは役に立たない可能性のあるフィールドのスクリーニング、必要に応じた新しい属性の取得、詳細なスクリーニングおよびサンプリング手法を使用したパフォーマンスの向上などを行い、フォローアップすることができます。

エンティティ分析と自動データ準備を組み合わせることによって、できる限りクリーンなデータを扱うことができます。

# 注意事項

この情報は、世界各国で提供される製品およびサービス向けに作成されています。

IBMはこのドキュメントで説明する製品、サービス、機能は他の国では提供していない場合があります。現在お住まいの地域で利用可能な製品、サービス、および、情報については、お近くの IBM の担当者にお問い合わせください。IBM 製品、プログラム、またはサービスに対する参照は、IBM 製品、プログラム、またはサービスのみが使用することができることを説明したり意味するものではありません。IBM の知的所有権を侵害しない機能的に同等の製品、プログラム、またはサービスを代わりに使用することができます。ただし、IBM 以外の製品、プログラム、またはサービスの動作を評価および確認するのはユーザーの責任によるものです。

IBMは、本ドキュメントに記載されている内容に関し、特許または特許出願中の可能性があります。本ドキュメントの提供によって、これらの特許に関するいかなる権利も使用者に付与するものではありません。ライセンスのお問い合わせは、書面にて、下記住所に送ることができます。

IBM Director of Licensing, IBM Corporation, North Castle Drive,  
Armonk, NY 10504-1785, U. S. A.

2 バイト文字セット (DBCS) 情報についてのライセンスに関するお問い合わせは、お住まいの国の IBM Intellectual Property Department に連絡するか、書面にて下記宛先にお送りください。

神奈川県大和市下鶴間1623番14号 日本アイ・ビー・エム株式会社 法務・知的財産 知的財産権ライセンス渉外

**以下の条項は、イギリスまたはこのような条項が法律に反する他の国では適用されません。** International Business Machines は、明示的または黙示的に関わらず、第三者の権利の侵害しない、商品性または特定の目的に対する適合性の暗黙の保証を含むがこれに限定されない、いかなる保証なく、本出版物を「そのまま」提供します一部の州では、特定の取引の明示的または暗示的な保証の免責を許可していないため、この文が適用されない場合があります。

この情報には、技術的に不適切な記述や誤植を含む場合があります。情報については変更が定期的に行われます。これらの変更は本書の新版に追加されます。IBM は、本書に記載されている製品およびプログラムについて、事前の告知なくいつでも改善および変更を行う場合があります。



IBM 以外の Web サイトに対するこの情報内のすべての参照は、便宜上提供されているものであり、決してそれらの Web サイトを推奨するものではありません。これらの Web サイトの資料はこの IBM 製品の資料に含まれるものではなく、これらの Web サイトの使用はお客様の責任によるものとします。

IBM はお客様に対する一切の義務を負うことなく、自ら適切と考える方法で、情報を使用または配布することができるものとします。

本プログラムのライセンス取得者が (i) 別途作成されたプログラムと他のプログラム（本プログラムを含む）との間の情報交換および (ii) 交換された情報の相互利用を目的とした本プログラムに関する情報の所有を希望する場合、下記住所にお問い合わせください。

IBM Software Group, Attention:Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA.

上記のような情報は、該当する条項および条件に従い、有料で利用できるものとします。

本ドキュメントに記載されている許可されたプログラムおよびそのプログラムに使用できるすべてのライセンス認証された資料は、IBM Customer Agreement、IBM International Program License Agreement、および当社とかわした同等の契約の条件に基づき、IBM によって提供されます。

ここに記載されているパフォーマンスデータは、すべて管理環境下で確認されたものです。そのため、他の操作環境で得られた結果は大きく異なる可能性があります。開発レベルのシステムで測定が行われている場合があり、これらの測定値は一般に利用可能なシステムと同じであることを保証するものではありません。また、測定値が推定値である可能性があり、実際の結果は異なる場合があります。本ドキュメントのユーザーは、特定の環境に適したデータを検証する必要があります。

IBM 以外の製品に関する情報は、それらの製品の供給業者、公開済みの発表、または公開で使用できるソースから取得しています。IBM は、それらの製品のテストは行っておらず、IBM 以外の製品に関連する性能、互換性、またはその他の要求については確認できません。IBM 以外の製品の性能に関する質問は、それらの製品の供給業者に通知する必要があります。

IBM の将来の方向性または意向に関する記述については、予告なく変更または取り消すことがあり、目的や目標のみを示すものです。

この情報には、日常の業務処理で用いられるデータや報告書の例が含まれています。できる限り詳細に説明するため、例には、個人、企業、ブランド、製品などの名前が使用されています。これらの名称はすべて架空のものであり、実際の企業で使用される名称および住所とは一切関係ありません。

この情報をソフトコピーでご覧になっている場合は、写真やカラーのイラストが表示されない場合があります。

## 商標

IBM、IBM ロゴ、および [ibm.com](http://ibm.com)、SPSS は、世界の多くの国で登録された IBM Corporation の商標です。IBM の商標の現在のリストは、<http://www.ibm.com/legal/copytrade.shtml> を参照してください。

Intel、Intel のロゴ、Intel Inside、Intel Inside のロゴ、Intel Centrino、Intel Centrino のロゴ、Celeron、Intel Xeon、Intel SpeedStep、Itanium、および Pentium は、米国およびその他の国の Intel Corporation または関連会社の商標または登録商標です。

Linux は、米国およびその他の国における Linus Torvalds の登録商標です。

Microsoft、Windows、Windows NT、および Windows のロゴは、米国およびその他の国における Microsoft 社の商標です。

UNIX は、米国およびその他の国における The Open Group の登録商標です。

Java およびすべての Java ベースの商標およびロゴは、米国およびその他の国の Sun Microsystems, Inc. の商標です。

その他の製品名およびサービス名等は、IBM または他の会社の商標です。



# 索引

- 削除
  - エンティティ リポジトリ, 49-50
- 商標, 74
- 消去
  - エンティティ リポジトリ, 48
- 設定
  - エンティティ リポジトリ, 25, 35
- nodes
  - エンティティ分析ストリームへの追加, 38
- エクスポート
  - エンティティ リポジトリへのデータ, 9
- エクスポート ノード
  - エンティティ分析, 9
- エンティティ分析
  - 定義, 1
  - IBM SPSS Modeler での使用, 4
  - 予測分析との比較, 3
- エンティティ タイプ
  - エンティティ分析, 31
  - エンティティ リポジトリ, 21
- エンティティ リポジトリ, 15
  - 作成, 6, 17-18
  - 削除, 49-50
  - 消去, 48
  - 維持, 26
  - 設定, 15, 25, 35
  - IBM SPSS Modeler への接続, 7
  - オプション, 19
  - 管理タスク, 44
  - フィーチャ, 29
  - ポート割り当ての設定, 45
- エンティティの一致、しきい値の設定, 34
- 識別した解決、分析, 36
- ソース タグ
  - エンティティ リポジトリ, 21
- 使用タイプ、エンティティ分析, 31
- データ ソース
  - エンティティ分析との接続, 5, 16
  - エンティティ分析の参照, 19, 42
- データ ソース、エンティティ分析の選択, 36
- データ型情報、エンティティ分析の設定, 37
- 法律に関する注意事項, 72
- 名前の変更
  - エンティティ分析のデータ フィールド, 37
- 識別の解決、エンティティ分析, 9
- 一意のキー
  - エンティティ分析, 7
  - エンティティ リポジトリ, 21
- 入力ノード
  - エンティティ分析, 11
- フィーチャ
  - エンティティ リポジトリ, 7, 20-21, 24, 26, 29, 41
- フィールドの関連付け
  - エンティティ リポジトリのフィーチャ, 7, 20-21, 24, 26, 41
- ポート割り当て
  - エンティティ分析との設定, 45
- リポジトリ
  - エンティティ分析, 6-7, 9, 15, 17-21, 24-26, 29, 31, 35, 41, 48-50
  - エンティティ分析の管理, 44