NTT DATA
Global IT Innovator

【OI】セマンティックデータソリューション　補足資料

２０１８年5月２５日
第四金融事業本部

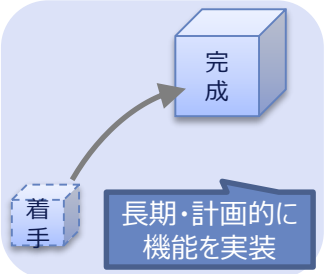
© 2018 NTT DATA Corporation

SoEシステムに求められること

**顧客のトップライン向上、顧客の顧客のCX向上に寄与する今後のSoEシステムには
データ活用できるまでの時間を極小化できる仕組みが必要。**

SoRシステム

開発手法

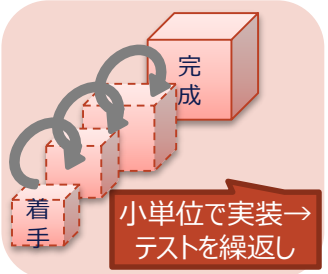


体制

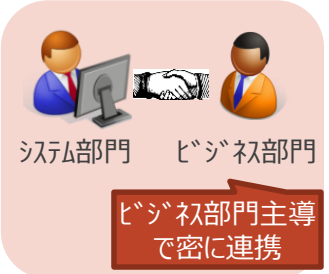
- **システム部門主導**で**ウォーターフォール**開発
- 新たなデータを追加する**都度モデリング**が発生
- 変更の都度、データ移行、リリース作業が発生するなど**変更に弱い**

SoEシステム

開発手法



体制



- 顧客接点の多い**ビジネス部門主導**で**アジャイル**でデータ整備し、APを開発
- **データの意味をシステムが理解**しモデリング不要
- **変更に強い**

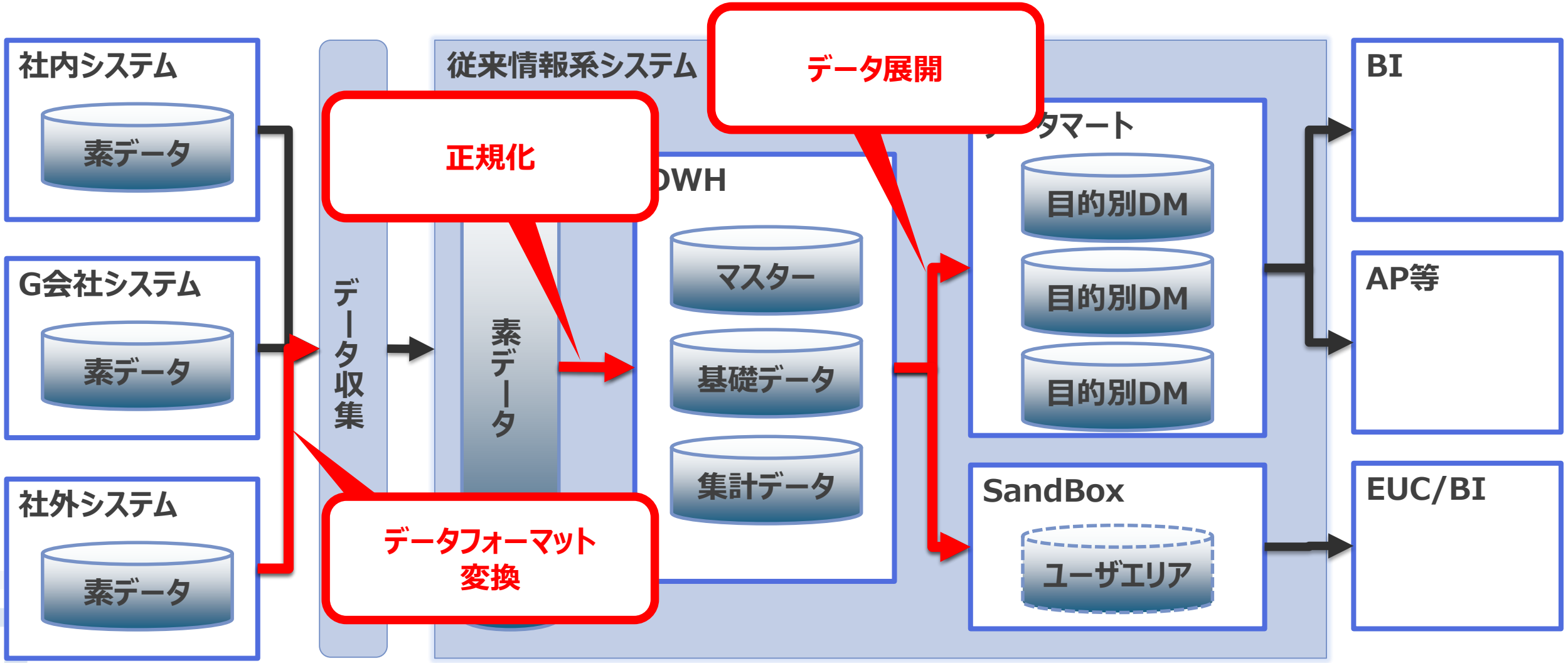
☑**顧客とのエンゲージメントを高める**モード２開発に取込んだデータを迅速に活用

© 2018 NTT DATA Corporation

NTT DATA

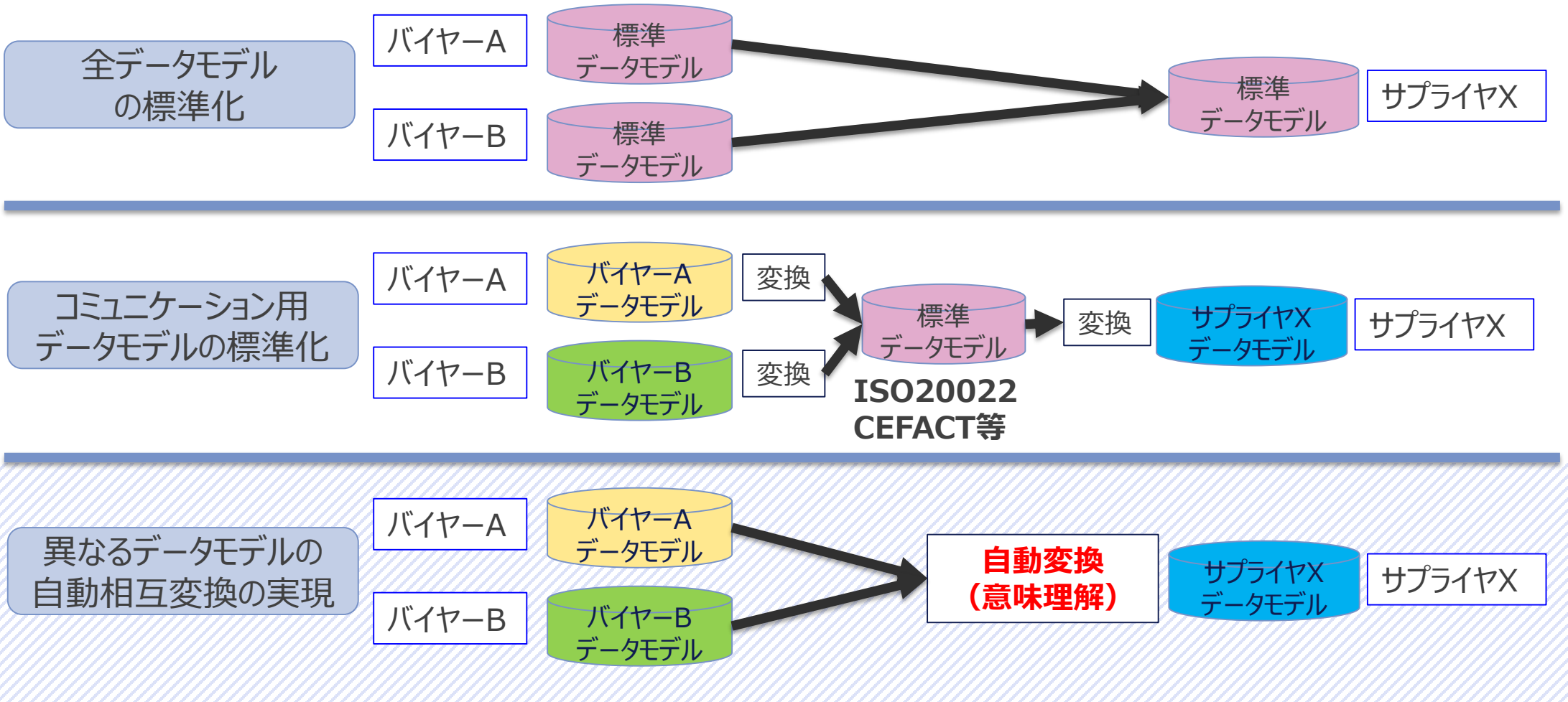
従来型手法におけるデータ整備の課題

従来のやり方ではいたる所で「標準化」が必要な仕組みとなっている。



データフォーマットの「標準化」の限界

各プレイヤーはそれぞれ異なるデータモデルや、フォーマットをもっている。
プレイヤー横断的な標準データフォーマット策定や、データモデル標準化（事象の捉え方の統一化）は限界。



各プレイヤーのデータモデル（ものの見方）を統一することは不可能

送受される情報が安定的でないと変更対応が困難。
事前に**厳密に標準を定めた情報しか送受できない**。

人間が必要な情報を含む**様式の異なる帳票を適切に理解するプロセスを、machine executeさせる**ことが有効

© 2018 NTT DATA Corporation

NTT DATA

セマンティックデータソリューションで実現したいこと

「デジタル社会における情報活用」においての実現したいことは以下の通り。

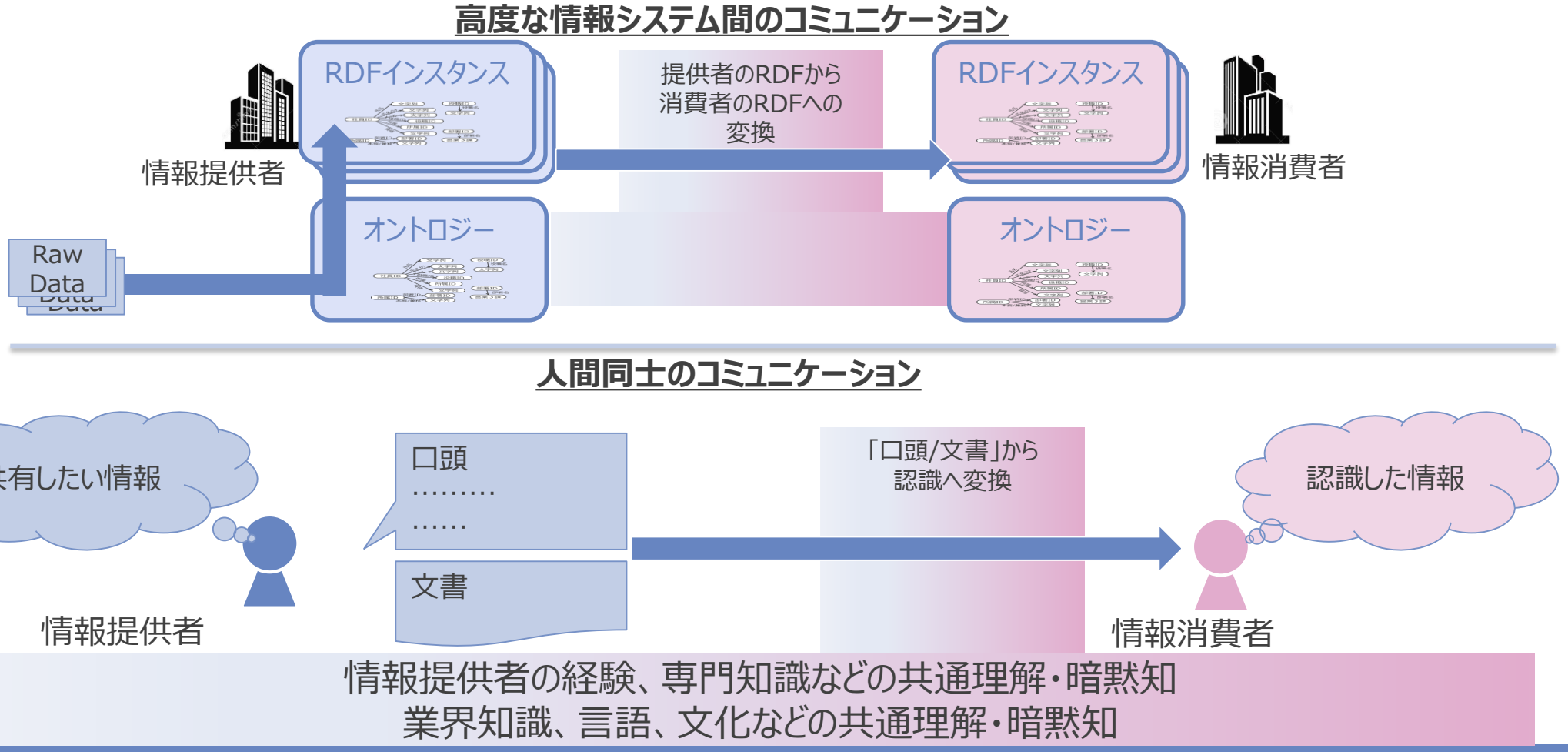
#	実現したいこと	詳細
1	**新しいデータが次々と発生しても、すぐに活用できる仕組み**	• 入ってきた新しいデータの「意味」を捉えてすぐに活用する ⇒ データモデリングという標準化作業を不要に
2	**異なるデータモデルの自動相互変換の実現**	• データにラベル（XMLタグ）と関係性（RDF）と意味理解（オントロジー）を付加することで実現 • 時間や負荷がかかるマッピング作業が不要となる

上記課題を解決する新技術の獲得が必要である。

© 2018 NTT DATA Corporation

オントロジーで相互理解する

個々の認識の仕方を統一すること（データモデルの標準化）は不可能だが、データだけでなく認識/理解も共有することで「違い」を吸収することが可能。これは人間同士のコミュニケーションと似通っている。



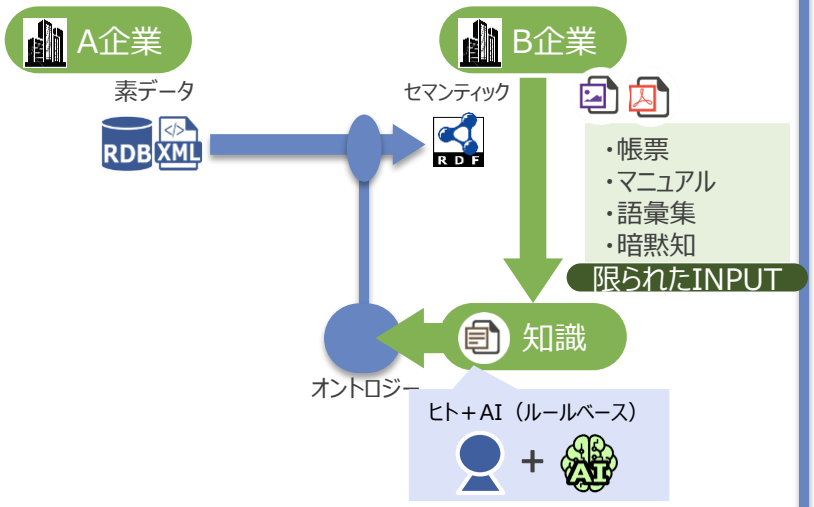
© 2018 NTT DATA Corporation

実現に向けてのステップ

1st Step

オントロジー	「ヒト＋AI(ルール)」で実現
元データ	素データ(ラベル＋データ)

元データを、自己オントロジーを用いて利用データ（セマンティック）へ変換して利用ができる。



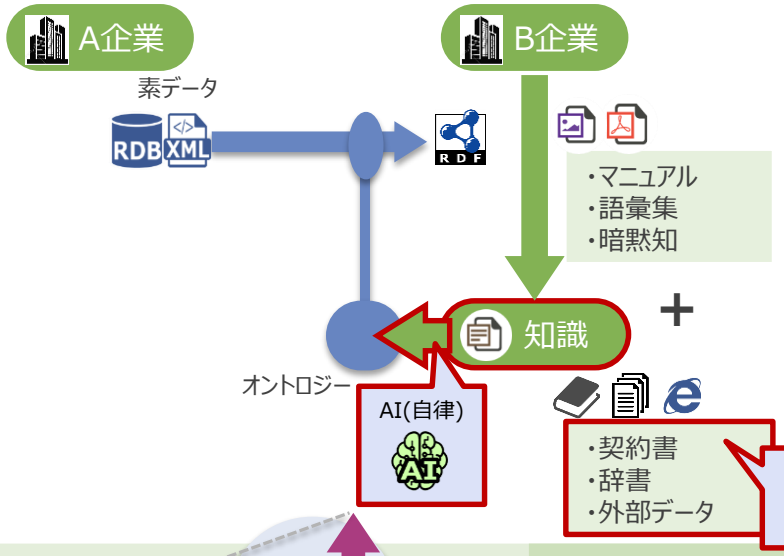
チャレンジ

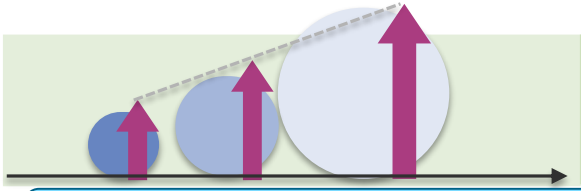
オントロジーと実データからセマンティックを生成

2nd Step

オントロジー	「AI」で実現
元データ	素データ(ラベル＋データ)

オントロジーを自律的に学習するため、高精度で変化に強い、データ活用が実現できる。





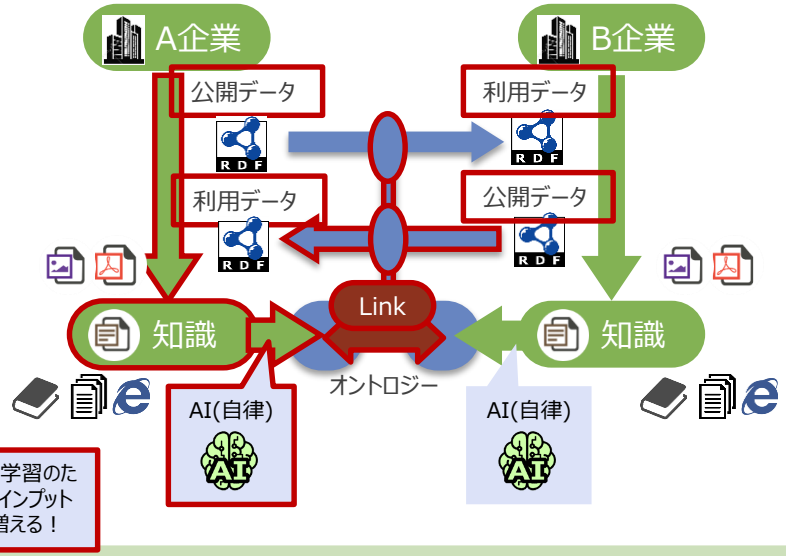
2nd Step以降はAIによりオントロジーが**自律的に成長**していく。

オントロジーが自律的に成長する

3rd Step

オントロジー	「AI」で実現
元データ	オントロジー＋RDFデータ

双方向でオントロジーを生成・リンクさせることにより、データを透過させる。(双方向のデータ活用)



オントロジー同士が意味をリンクする

技術要素の獲得状況

#	技術要素	詳細	当社における状況
1	オントロジーとセマンティックの実装技術	• オントロジーはOWL、セマンティックはRDFでの実装が主流 • 問合わせ言語はSPARQL、GRAPHQLが主流 • 非機能面、運用面での課題はこれから	• RDFをMarkLogicに格納し、SPARQLで問合わせを行うことは**実施済み** • オントロジー実装は机上にて実施済
2	オントロジー構築時におけるモデリング技法	• 専業コンサルタントもいくつか出てきており、モデリング技法について確立されつつある	• 未着手。**顧客とのPOCを通じて獲得したい**
3	オントロジーと実データをマッピングしセマンティックを自動生成（エンティティリンキング）	• 学術研究レベル • Wells Fargo、Franklin Templetonがパイオニア。	• **FY17にMD研と共同研究**を実施
4	リファレンスオントロジーを自動拡張	• 未着手	• 未着手
5	オントロジー間をリンクする	• 未着手	• 未着手

© 2018 NTT DATA Corporation

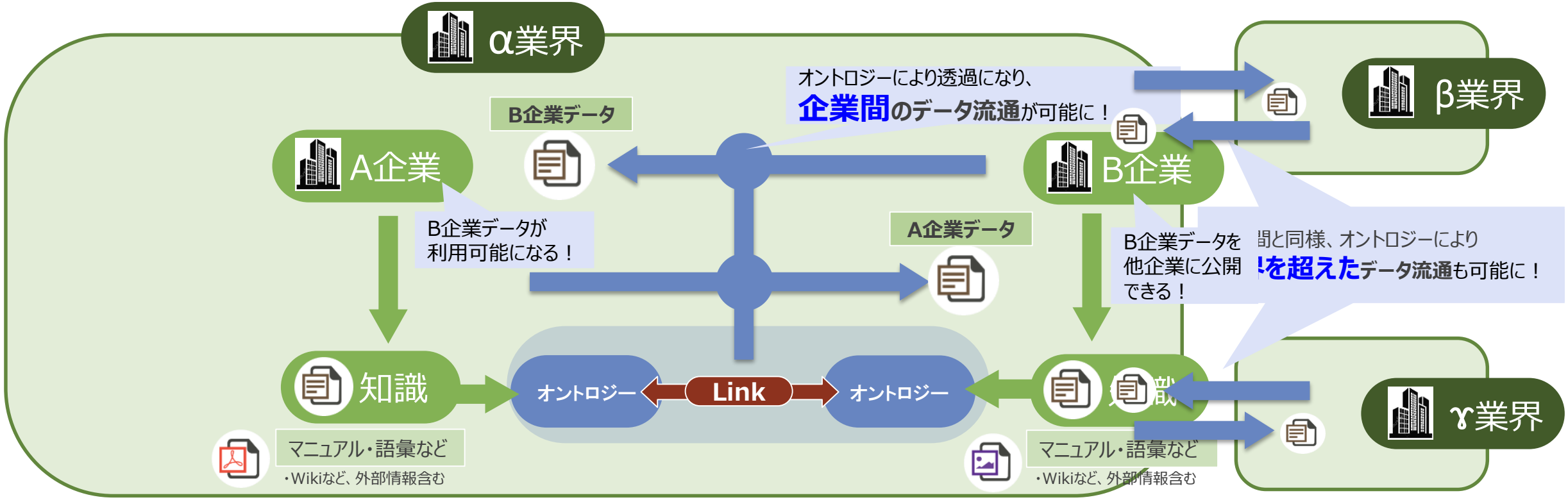
【R&D】実現のためのステップ

	1st Step		2nd Step以降
	FY2018 2Q	FY2018 3～4Q	FY2019 1Q～
対応	オントロジー構築技法の習得	オントロジーを参照し、 実データをセマンティックに マッピング	オントロジーの自動拡張 オントロジー間リンキング
確認事項 成果物	• OWLでオントロジーの構築 • AI（ルールベース）の実装 • OWL構築ツール選定 • 課題の認識	• オントロジーを参照し、実データをセマンティック変換するプロトタイプ開発 • 課題の認識 • 各課題に対するタスク設定	• オントロジーで理解できなかったデータの訂正方法の確立 • オントロジーが自律的に拡張するための検討及びプロトタイピング • 課題の認識

© 2018 NTT DATA Corporation

目指す世界 – データ流通革命 –

RDF・オントロジーベースのセマンティックな世界で、企業・業界を超えた**デジタル情報活用を実現**



#	データ流通の実現	データ爆発への対応
セマンティックな世界	○ 業界A ⟷ 業界B スキーマレスのため、データ流通との**親和性高**	○ 新データに対応するセマンティックの**自動構築**

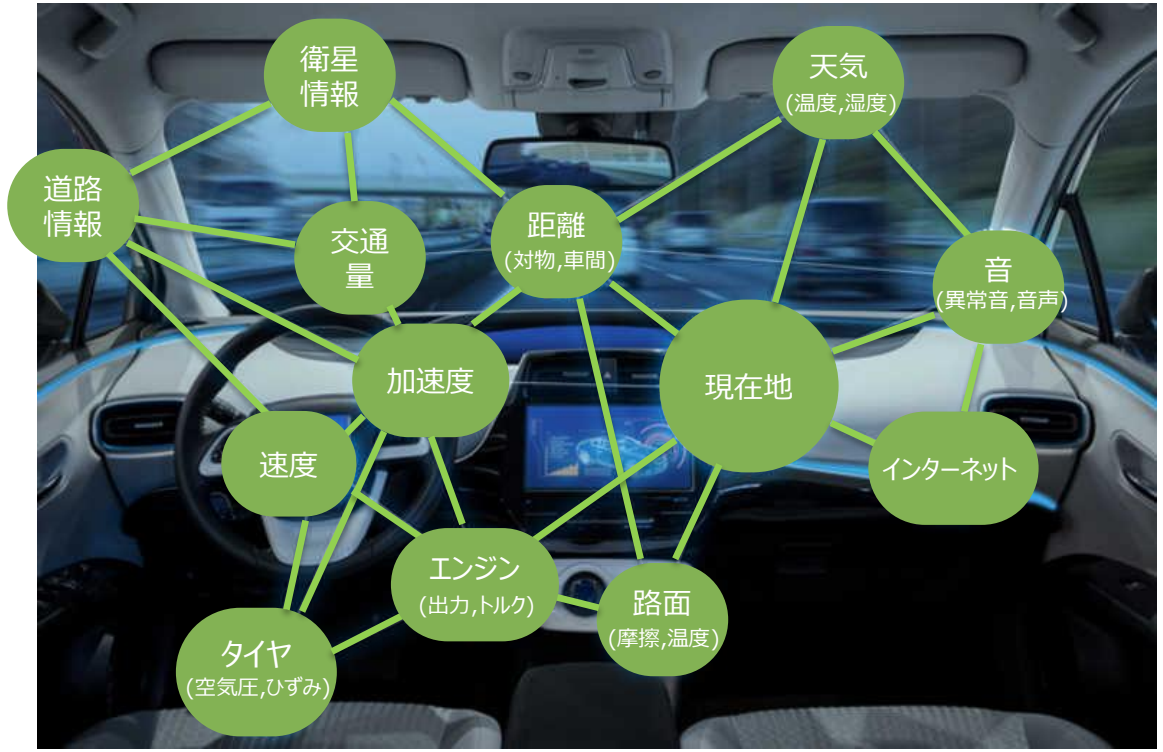
目指す世界 － 新規データの即活用 －

新規データが発生しても、オントロジーを用いることで**データの意味を捉えてすぐに活用**が可能。

インプット・分析

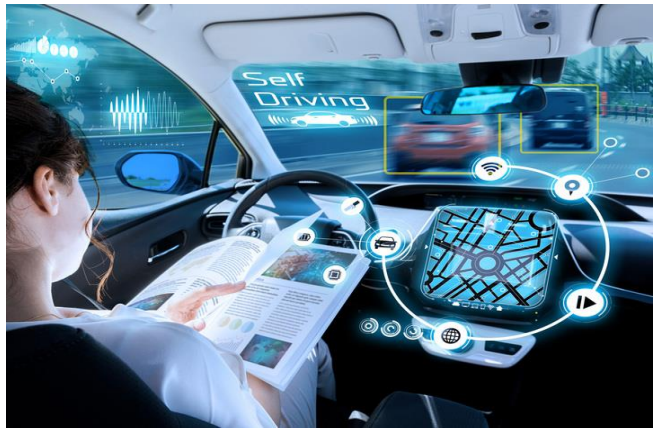
オントロジーを用いた多様・親和性高なデータ群

利用可能なデータは即活用



アウトプット

AIの自律的な動作



例えば、IOTから収集できるさまざまなデータを次から次に活用していくことで、より確実な予測・分析モデルを構築することが可能になる。

© 2018 NTT DATA Corporation

【ご参考】オントロジーで定義されるものとは

オントロジーでは機械が処理を行えるようにデータに意味を持たせる

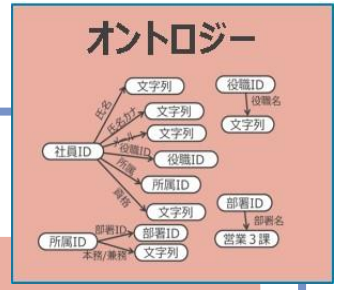
Human Executableな世界

マニュアル

辞書

暗黙知

Machine Executableな世界



データの意味

- ラベル
- 説明文
- リファレンス先
- 保持する属性

表現のルール

- 必須項目
- データ型
- データ長
- コード値
- 範囲

データ間の関係

- 同義／反対
- 類似
- 親子関係
- 包含関係
- 集合関係

由来・来歴

- 作成日
- 作成者
- 承認日
- 承認者
- 変更日
- 変更者
- バージョン
- ソース

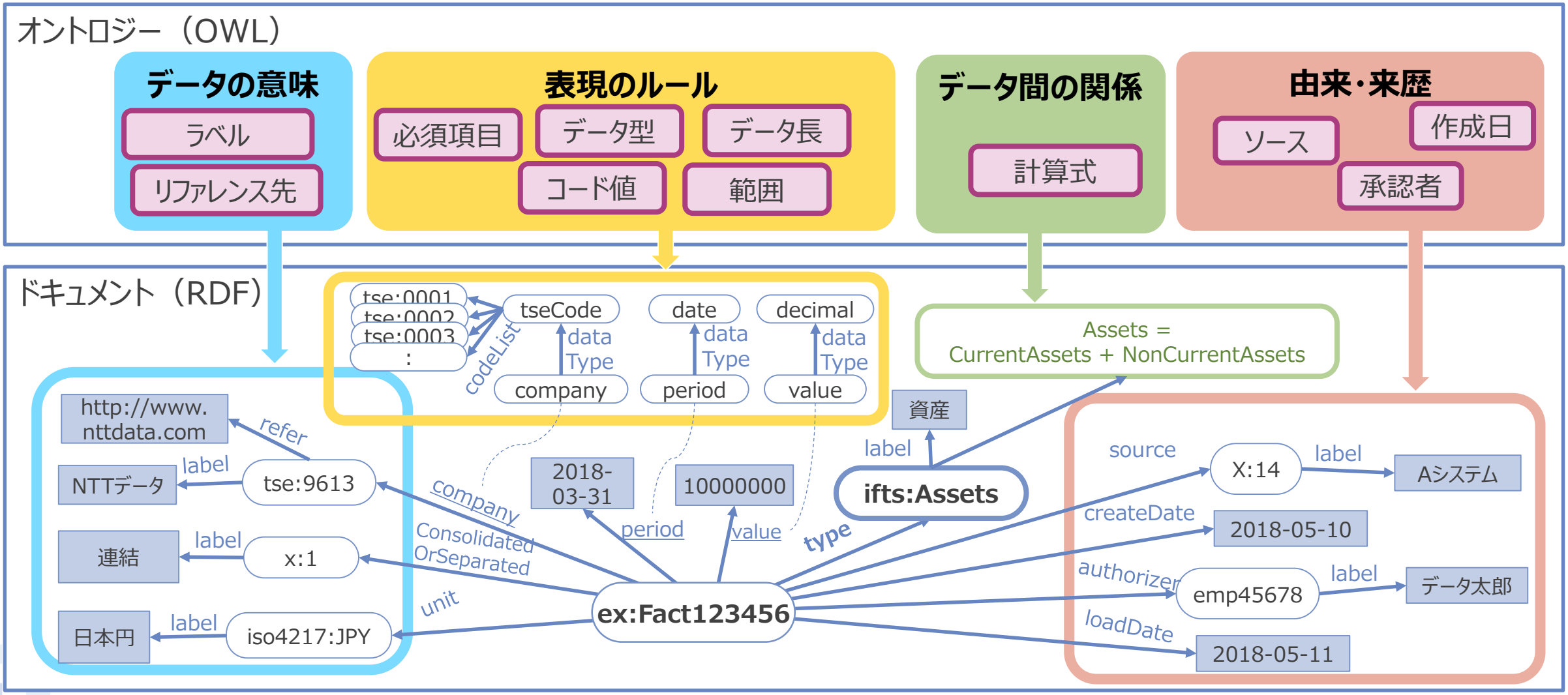
人間が業務を行う際に必要となる知識をオントロジーとして定義を行う

RDBの世界ではカラムを追加しないと意味や由来の情報を追加することはできない。

オントロジーでは柔軟に情報を追加することが可能（タグを追加するイメージ）

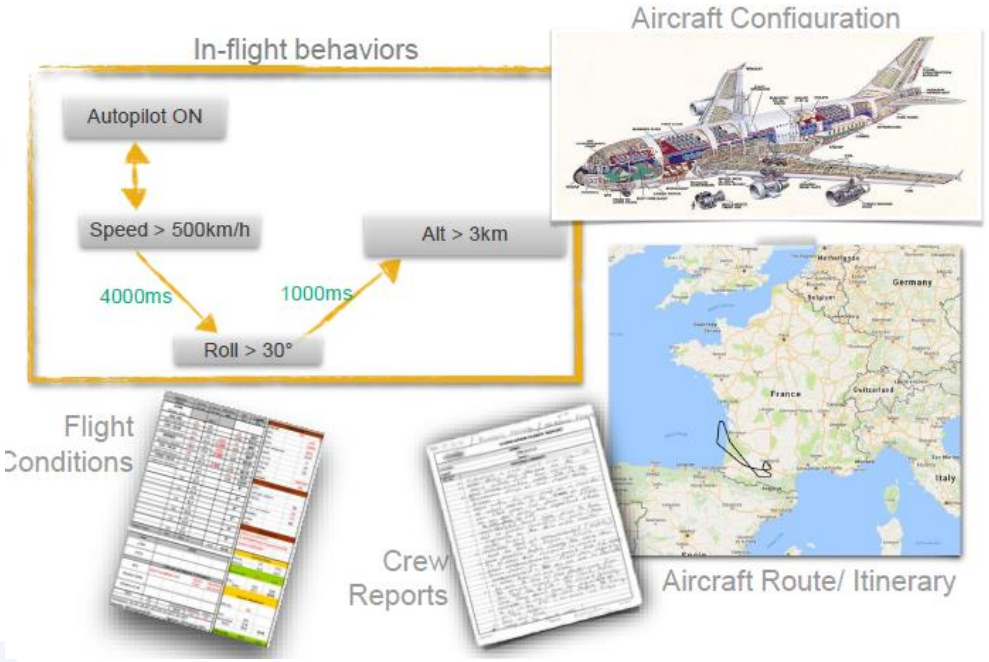
セマンティックとオントロジーの関係

オントロジーはモデルでありインスタンスはセマンティック（RDFドキュメント）で実装される。



IoTユースケース（新たに生まれ続ける分析ニーズ対応の重要性）

航空機には製造段階で大量のセンサーが組み込まれる。全ての分析観点を事前に定義することは不可能であり、生み出される情報をありのままに蓄積し、新たな分析ニーズに機動的に対応できることが必要。



- Sensor data
- Flight route
- Flight conditions
- Crew reports
- Crew conditions
- Passengers data
- Air controller data

⋮

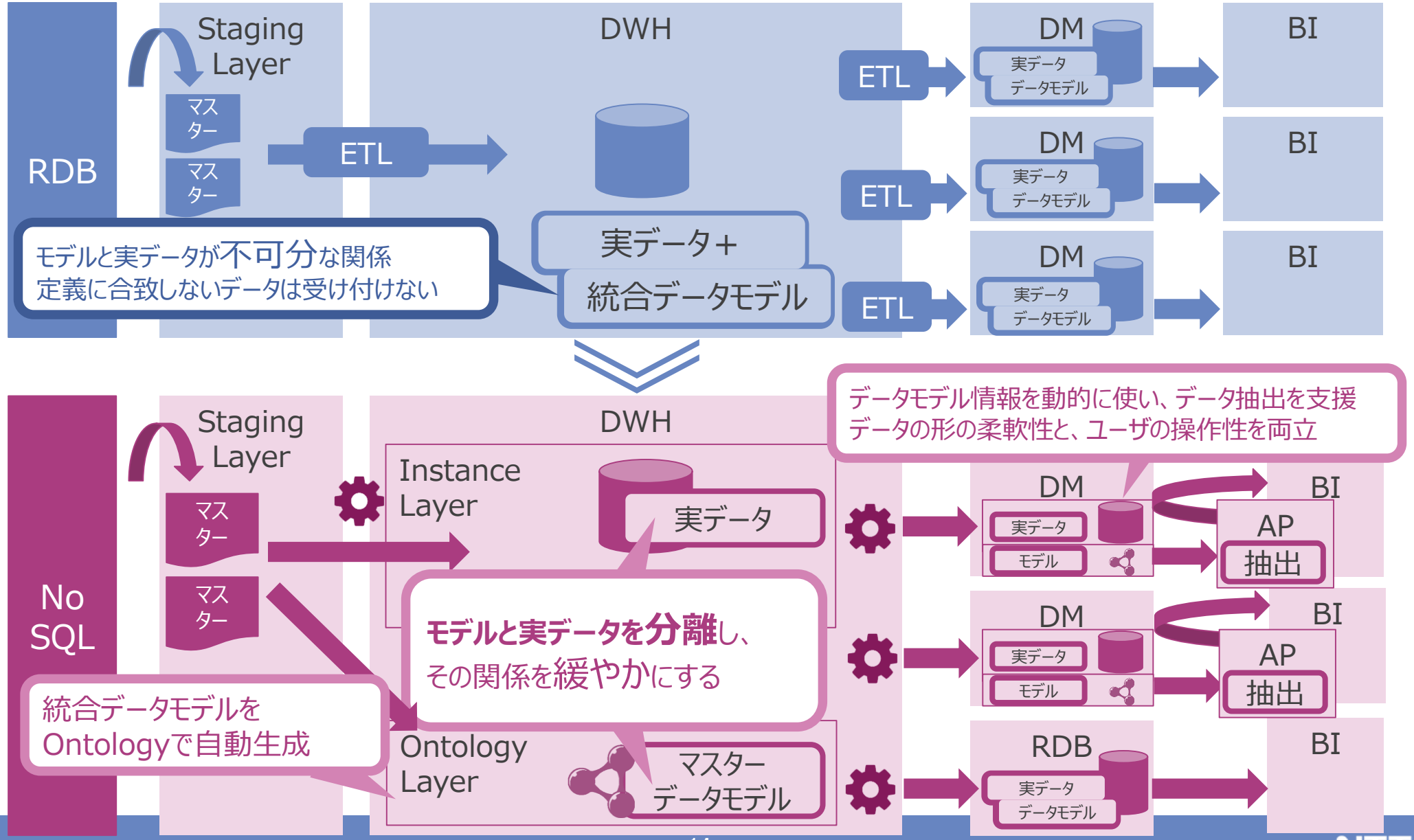
目的に応じた様々な分析

✓ 事故分析
今までにないニアミスについて気象条件が関係するか調べたい・・・

✓ 燃費等コスト最適化分析
航空機の揺れとリピート利用の相関を調べたい・・・

⋮

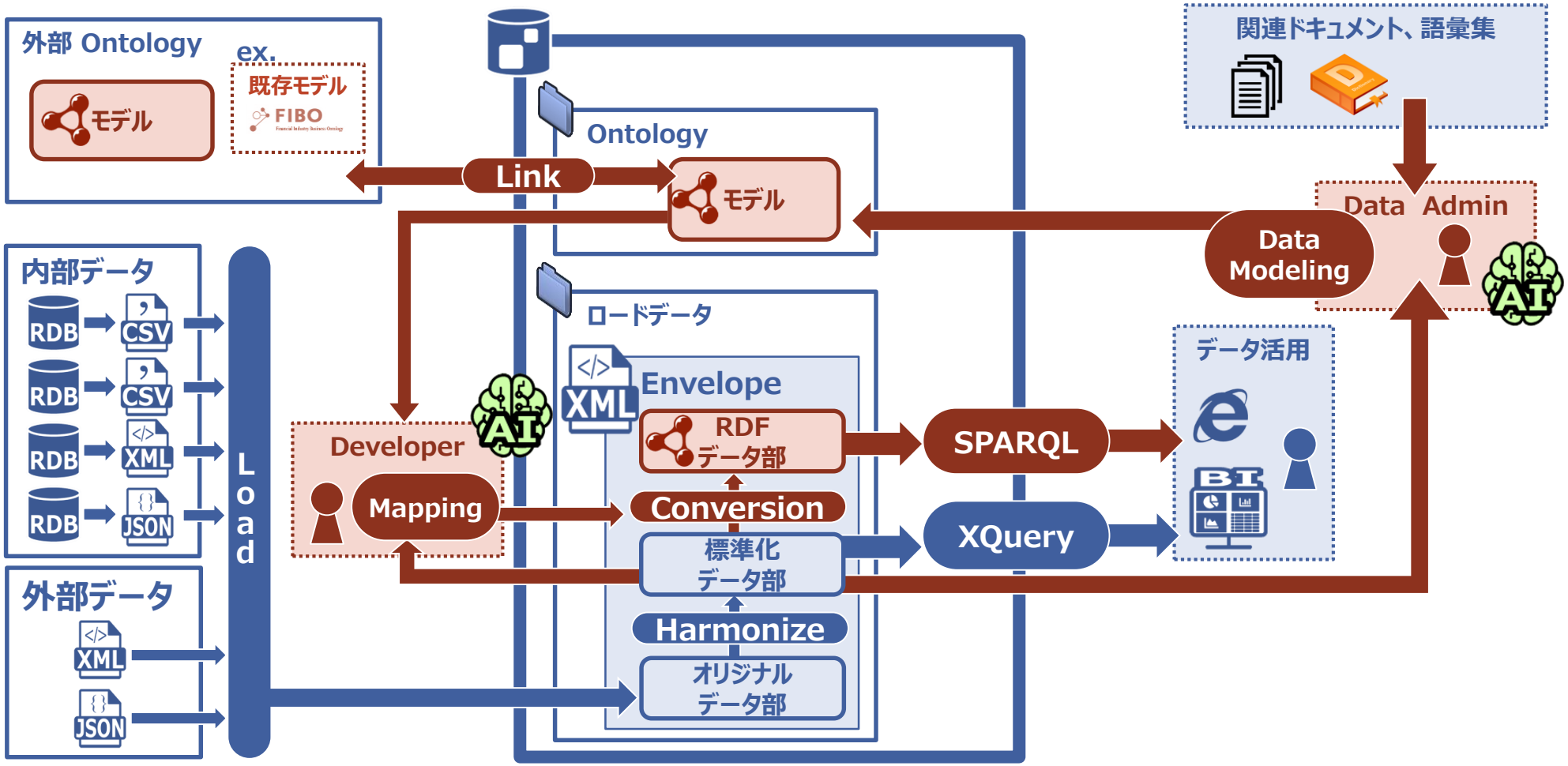
次世代MDMユースケース



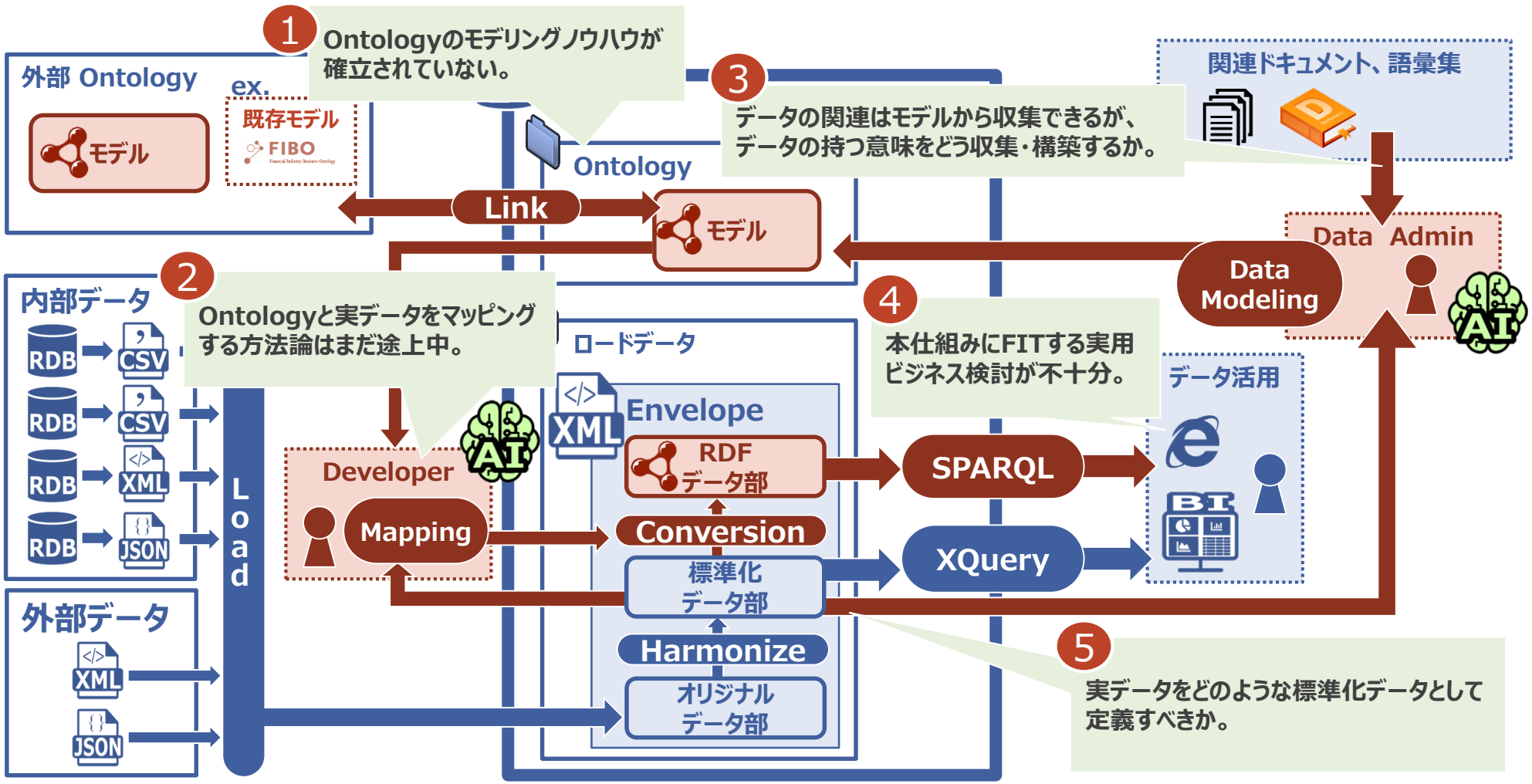
© 2018 NTT DATA Corporation

NTT DATA

【ご参考】Semantics 全体像



【ご参考】Semantics実装における課題



【ご参考】Semanticsに関する課題一覧

#	課題	詳細
1	**Ontologyのモデリングノウハウが確立されていない。**	・「社員」は「組織」に含まれます。という包含関係 ・「社員」は「社員コード」を持ちます。というドメイン情報 ・「社員コード」は「数字7桁」です。というレンジ情報 等の意味記述の**定義方法が確立されていないため、業務モデルを作成できない。**
		対象モデルを作成するにあたり、どのような観点で業務分析を行えばオントロジーを作成できるのかという**方法論が整理できていない**ため、**モデル構築を行っても、正しさを検証できない。**
2	**Ontologyと実データをマッピングする方法論はまだ途上中。**	オントロジーとデータを結ぶ**マッピング方法や、マッピング精度がまだ実用段階ではない。** 引続き、金マで検討する。
3	**データの関連はモデルから収集できるが、データの持つ意味をどう収集・構築するか。**	「契約者」が「契約金」と関係することは、関係モデル等から引合せることで抽出できる。 一方で、「契約」が「合意」という意味の狭義であるというような、語彙（イミ）を収集・定義することを標準化データ内から引合せることが困難。（標準化データに情報が存在しない可能性が高い。） ROWデータ内から**オントロジーを構築する方法・データを引合せるための補完情報の検討を行わないとオントロジーの自動構築が達成できない。**
4	**本仕組みにFITする実用ビジネス検討が不十分。**	RDFと親和性の高いビジネスの適用領域がまだ定まっていない。 RDFはセマンティック（意味論）のため、1と0の間を表現することに特化しており、ある問いに対して100%正解の答えが要求されるような厳密なシステムには親和性が低いと思われる。 これら**RDFの特性に当てはまるようなビジネス領域を見定めることができない。**
5	**実データをどのような標準化データとして定義すべきか。**	データ（インスタンス）とオントロジー（モデル）の境界はどこかを定義するための検討が必要。 データ/オントロジーの境界を定義しておかないと、**HarmonizeやAIMappingなどの役割境界を定めないと、実現に向けた検討内容がぶれてしまう。** 例えば、システム間における、性別の定義が「男性/女性」、「M:F」等で意味がずれている場合、これを吸収するレイヤは、Ontologyなのか、Harmonizeなのか、等の決めが必要。コード値等の正規化用途で構築された、機械のためのデータはどう扱うかについても同様の取り決めが必要。

© 2018 NTT DATA Corporation

NTT DATA

Global IT Innovator

© 2018 NTT DATA Corporation