

「今こそ求められるメタデータガバナンス」 一個別定義から統合へ向けて

2013年11月8日

株式会社データアーキテクト

真野 正

概要

- 企業内外に蓄積されたデータの山を「ゴミ」と化すか「宝」と化すかが、ビジネスの成否を左右する時代といっても過言ではない。
- 膨大かつ多種多様なデータの山から必要とする時に必要なデータが何処にあり、どのようにして取り出せば良いのかを教えてくれるのが、「メタデータ」である。
- 従来、「メタデータ管理」というと情報システム資産の開発あるいは保守時の影響・波及分析というIT部門内での効率化目的がほとんどだった。
- 今求められている「メタデータ」は、データの意味、変遷の履歴、データの所在、加工ルールといったデータを分析していく上での羅針盤(地図)となるものであり、かつ、データ利用者間での共通のナレッジとなるものである。
- 「メタデータ」はデータの品質保証書としてデータに添付されなければならない。
- データ活用のために避けて通ることができない「メタデータ」について意義を問い、それをどのような手順で整備・統制して行ったらよいか。

http://dataarch.co.jp/HPB17_DataArch/index.html

目次

1. メタデータの変遷
2. データ品質保証としてのメタデータ
3. メタデータ整備の手順
4. メタデータ統制

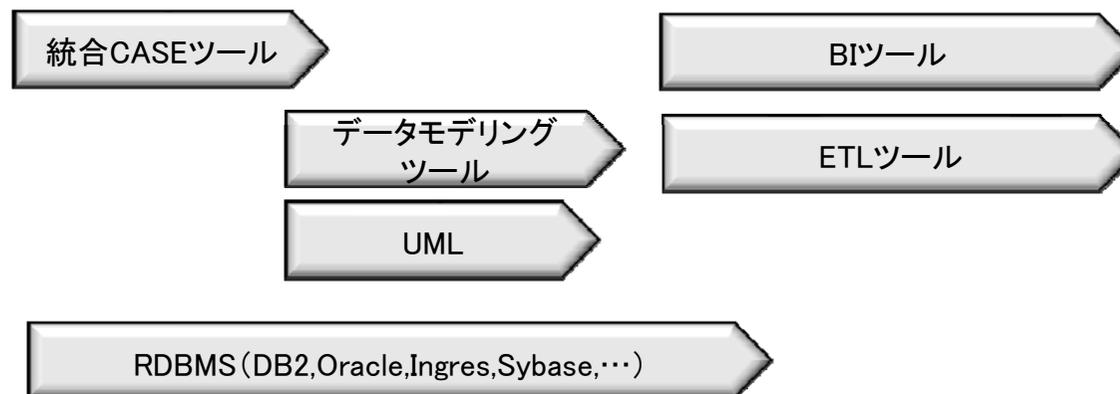
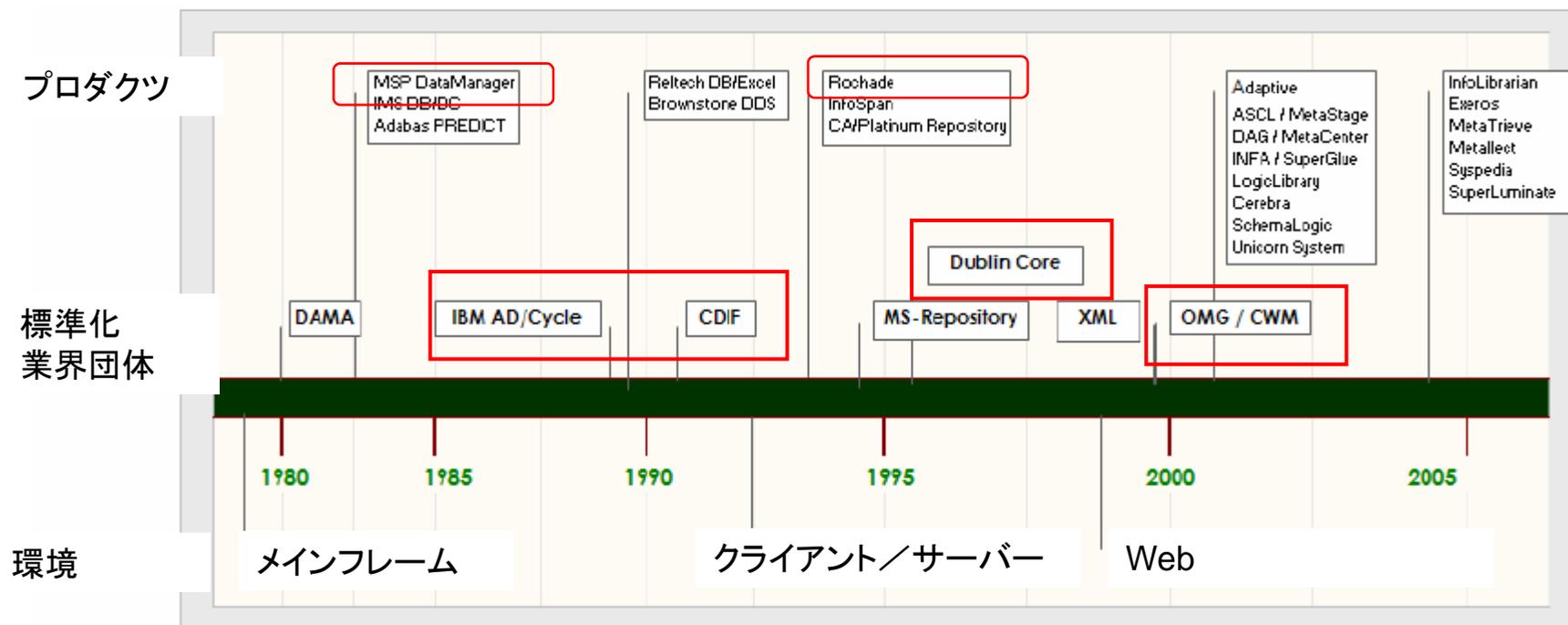
メタデータの変遷

1. 情報システム保守を主眼としたメタデータ管理
2. 管理対象の変遷: 定義、利用者、目的



メタデータリポジトリの歴史

- メタデータは、RDBMSのディクショナリ(システムカタログ)から始まった...



初期のメタデータ管理(80年代)

■ 情報システム開発・保守のためのメタデータ

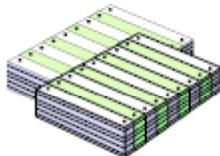
Mainframe Computers

- IBM
- Amdahl
- DEC
- HP



Program Source Code

- COBOL
- JCL
- Assembler



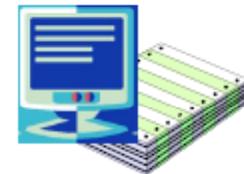
Records, Files, Copybooks

- VSAM / ISAM
- COBOL



Screens, Forms, & Reports

- CICS
- FOCUS
- Easytrieve



DBMS schemas

- IMS
- Adabas
- IDMS
- Ingres, Supra, DB2



Human Knowledge

- Documentation
- Data Definitions
- Business Rules
- User Requirements
- Domain Experts

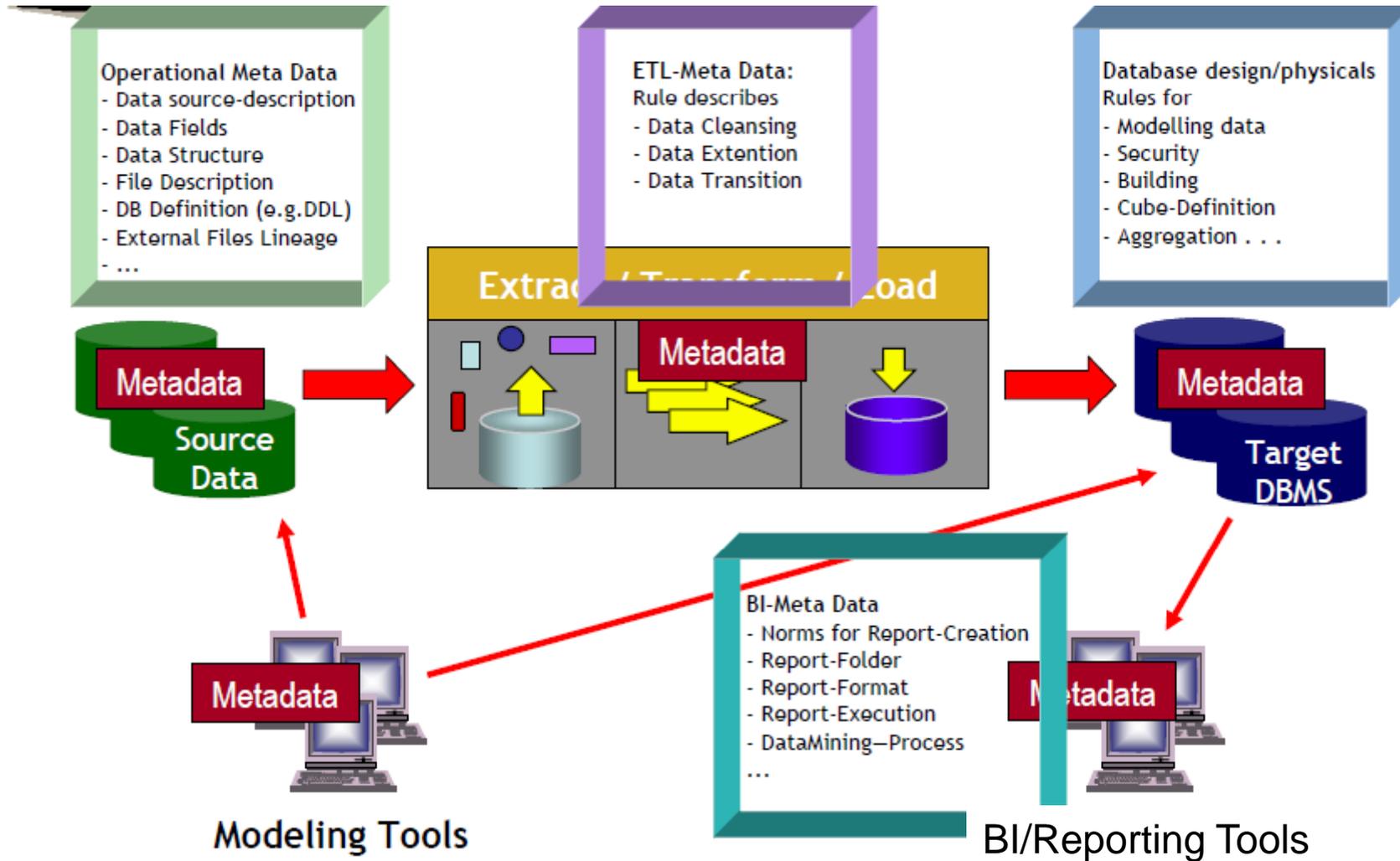


capturing "metadata"

- definitions
- descriptions
- inter-relationships
- "catalog" of corporate information assets

ツール毎のメタデータ管理

- ツールの数だけメタデータが存在する

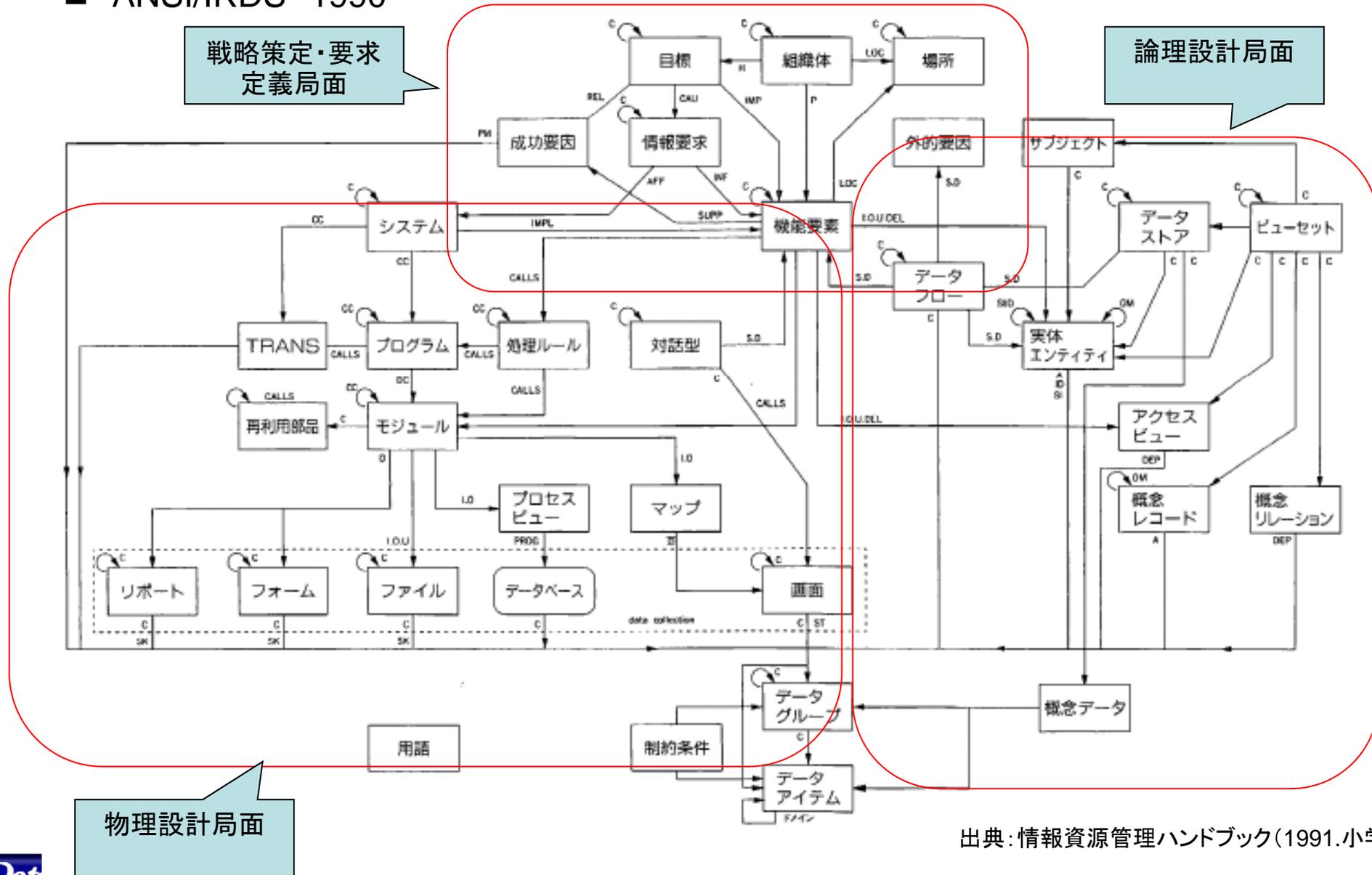


メタデータ標準化

- IRD: Information Resource Dictionary
 - ◆ ANSI/IRDS 1990
- CWM: Common Warehouse Model
 - ◆ OMG.MOFベース
- ISO11179
- DuburinCore

IRD (: Information Resource Dictionary) 標準メタ構造

■ ANSI/IRDS 1990



出典: 情報資源管理ハンドブック(1991.小学館)

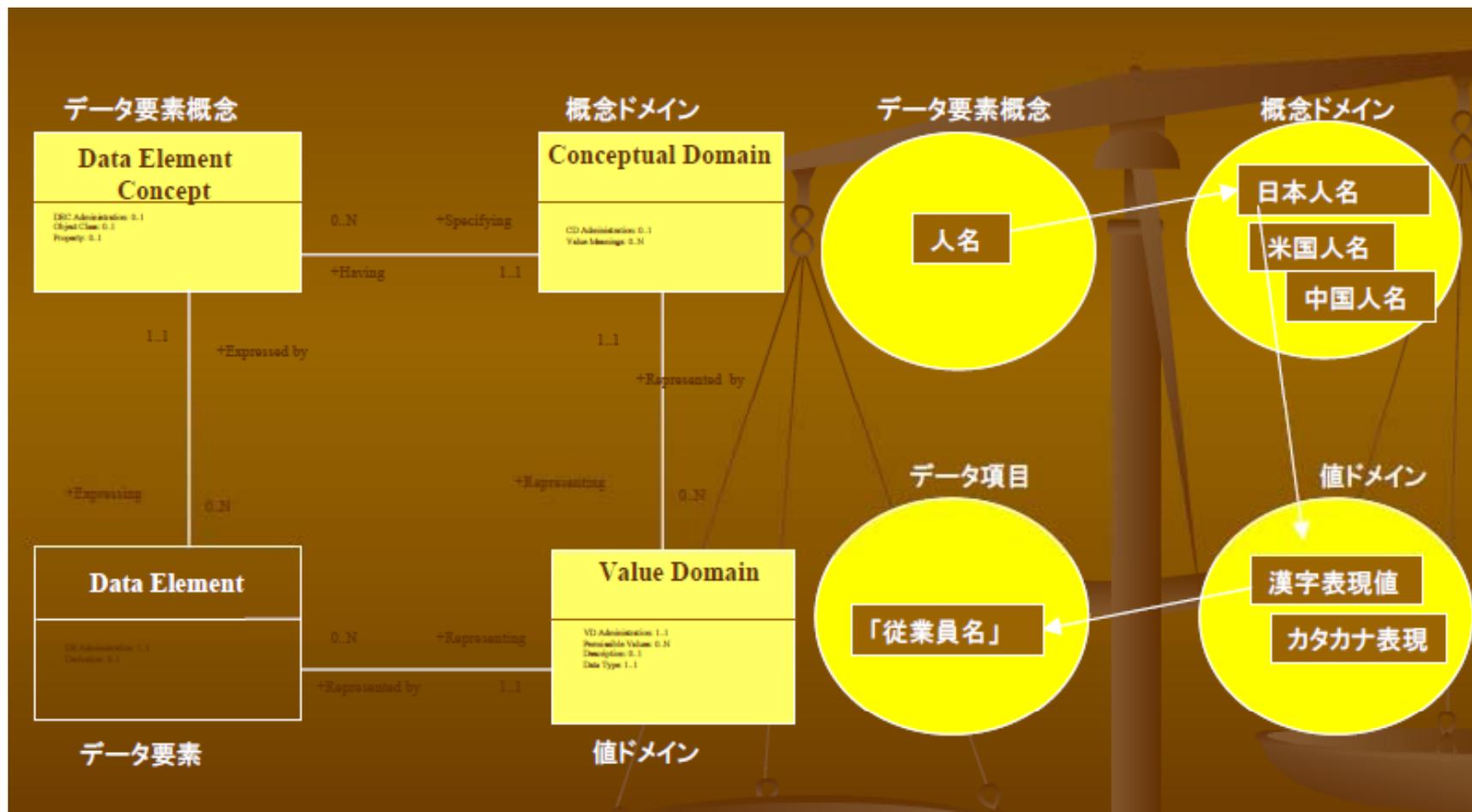
CWM (Common Warehouse Model)



管理	ウェアハウスプロセス			ウェアハウスオペレーション		
分析	変換		OLAP	データ マイニング	情報可視化	ビジネス 用語体系
手段	オブジェクトモデル	リレーショナル	レコード	多次元		XML
基礎	ビジネス情報	データタイプ	式 (Expression)	キーと インデックス	タイプ マッピング	ソフトウェア 配備
オブジェクトモデル						

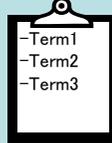
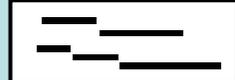
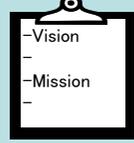
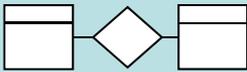
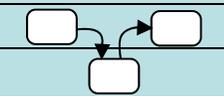
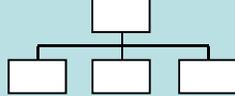
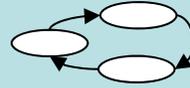
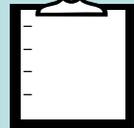
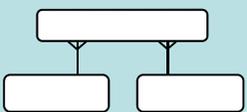
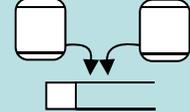
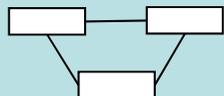
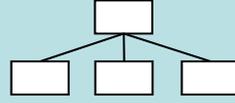
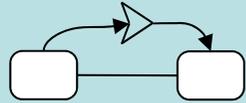
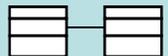
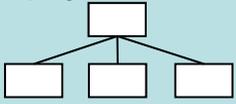
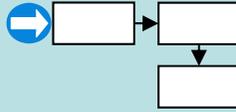
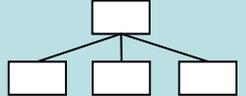
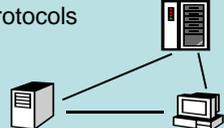
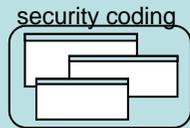
出典: Common Warehouse Metamodel
(CWM) Specification【OMG】

Metadata Registry Standards(ISO/IEC 11179)



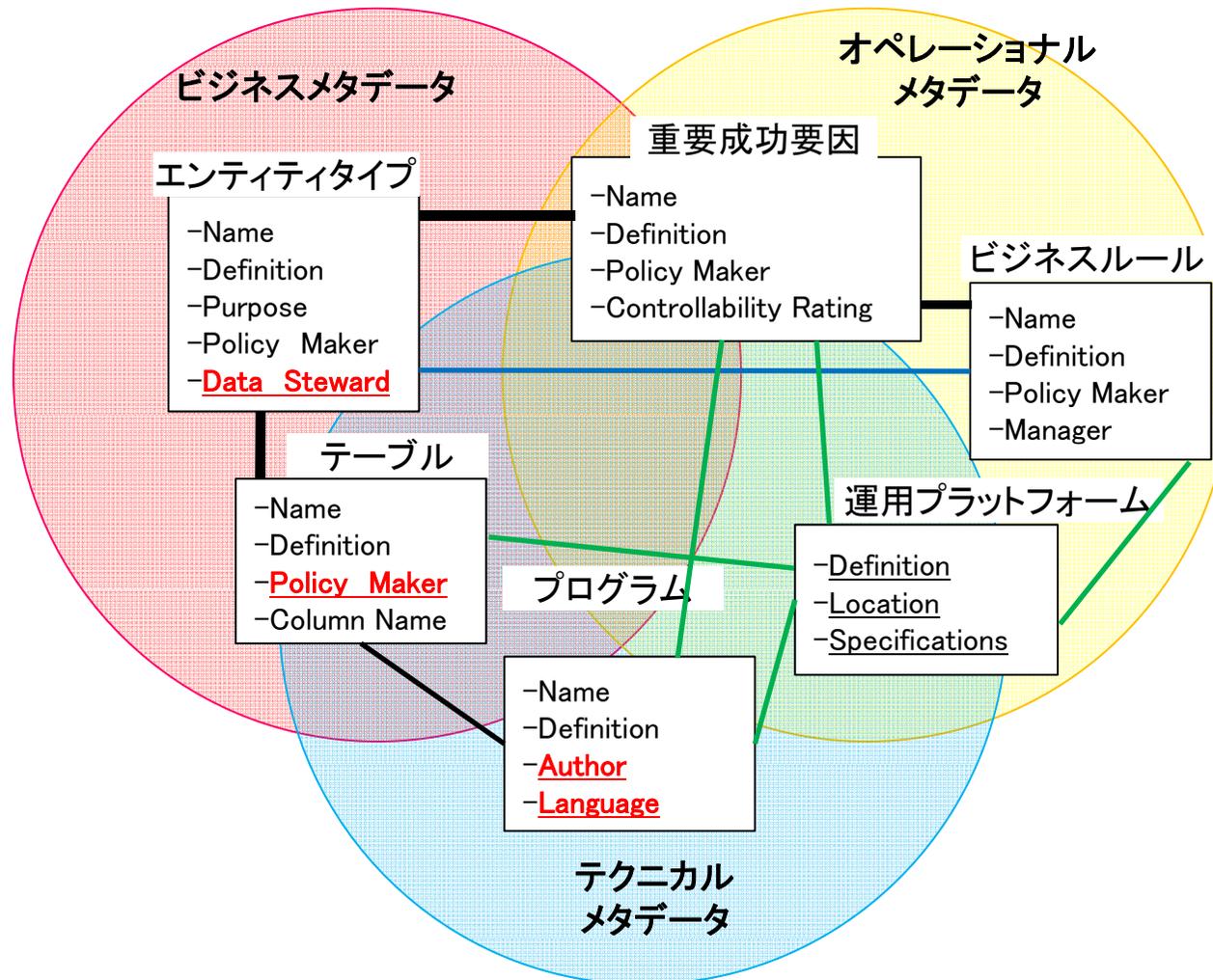
出典: 堀内一氏、標準化フォーラム2004発表資料

メタデータの管理対象

	Data (What)	Activities (How)	Locations (Where)	People (Who)	Time (When)	Motivation (Why) 
Objectives/ Scope (Planner's view)	List of things important to the enterprise 	List of functions the enterprise performs 	List of enterprise locations 	Organization approaches 	Business master schedule 	Business vision and mission 
Enterprise Model (Business Owner's View)	Language, divergent data model 	Business process model 	Logistics networks 	Organization chart 	State/transition diagram 	Business strategies, tactics, policies, rules 
Model of Fundamental Concepts (Architect's View)	Convergent e/r model 	Essential data flow diagram 	Locations of roles 	The viable system, use cases 	Entity Life History 	Business rule model 
Technology Model (Designer's View)	Database design 	System design, program structure 	Hardware, software distribution 	User interface, security design 	Event processing 	Business rule design 
Detailed Representation (Builder's View)	Physical storage design 	Detailed program design 	Network architecture, protocols 	Screens, security coding 	Timing definitions 	Rules specification program logic 
Functioning System	(Working System) Zachman Framework					
	Databases	Program inventory, logs	Communications facilities	Trained people	Business events	Enforced rules

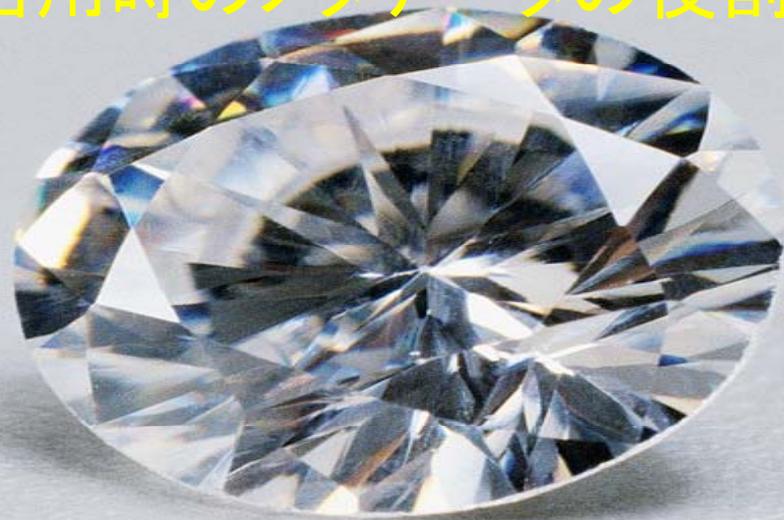
メタデータのタイプ

- メタデータは、その目的から3種類に分類できる

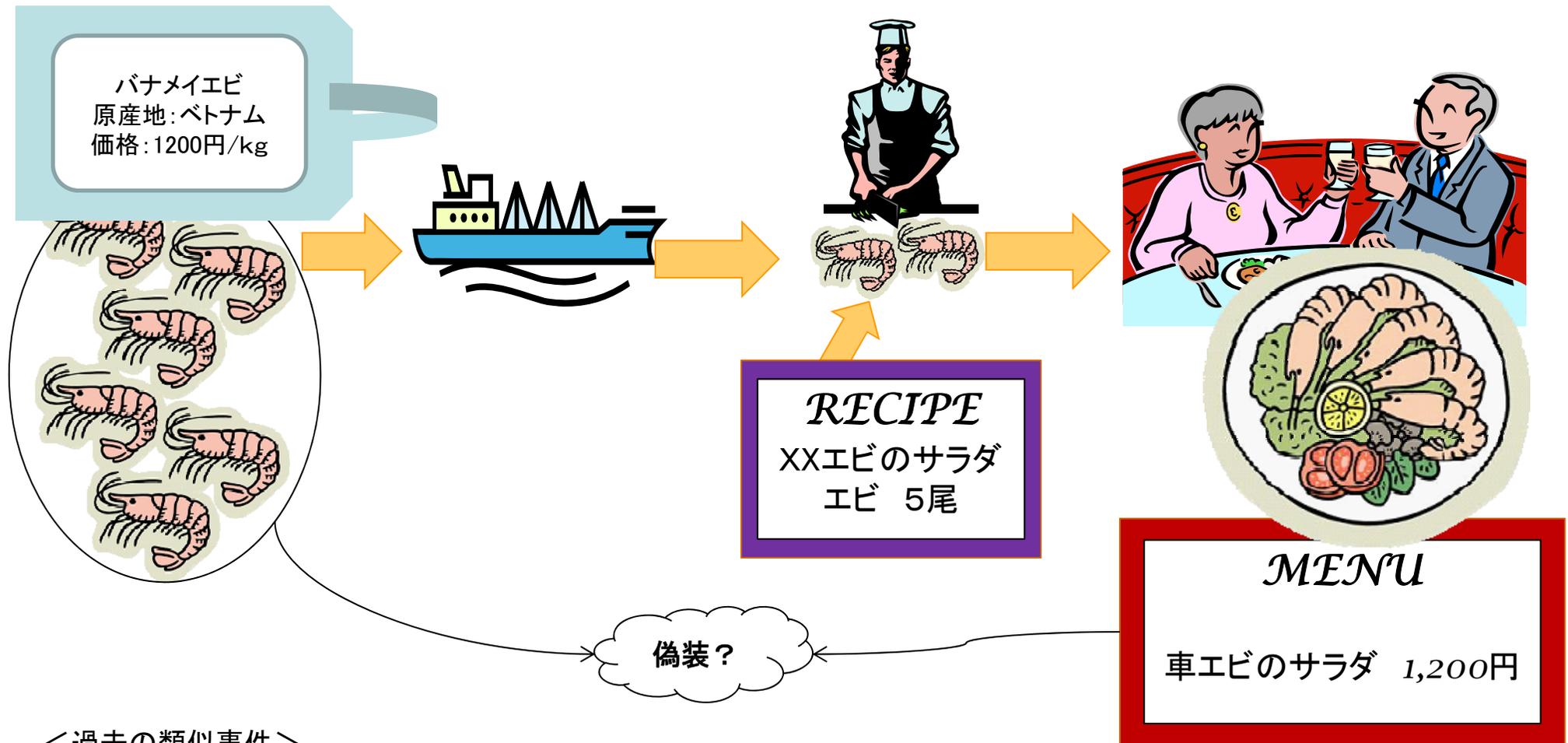


データ品質保証としてのメタデータ

1. データ活用時のメタデータの役割



食品偽表示



<過去の類似事件>

- 売れ残り商品を冷凍して保存、解凍した日付を製造日とする、賞味期限を本来の日付より延長して表示するなど、食品衛生法に違反していた問題。
- 中国で製造され、輸入された冷凍ギョーザを食べた人たちが中毒症状を起こし、検査の結果、ギョーザから農薬成分が検出された。

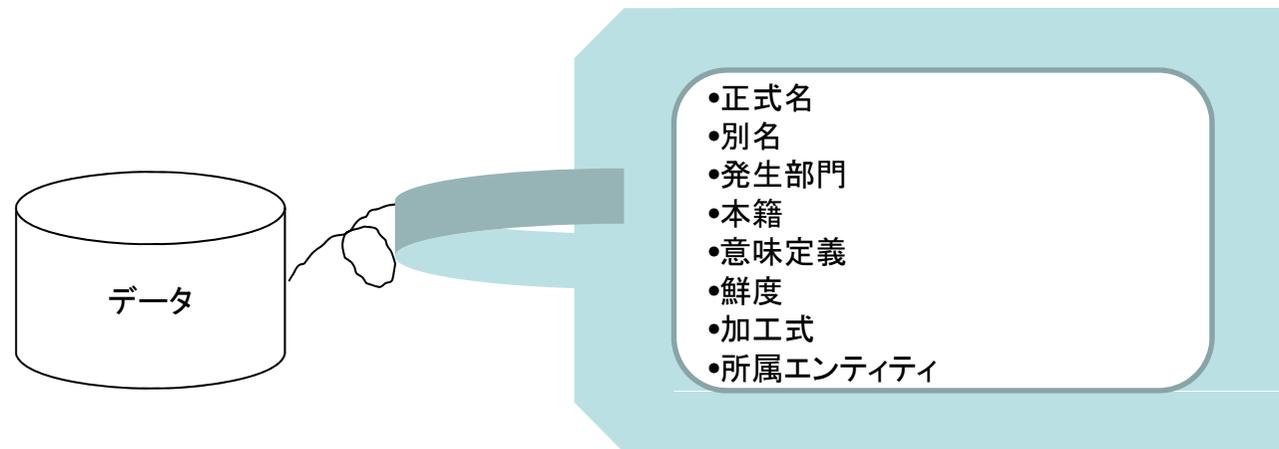
背景

- タグ付けによる産地保証
- 産地から最終消費者に至るトレーサビリティ
- 流通経路でのすり替え
- 商品は保証されているが人為的にすり替え
- 利用者側の知識不足

データ品質にも通じる

データの保証機関

- データの保証機関での証明付きのデータ
 - 証明付きの最低限の条件は、メタデータが添付されていること
- データ監理部門：社内データの保証
- データ保証機関：社外流通データの保証

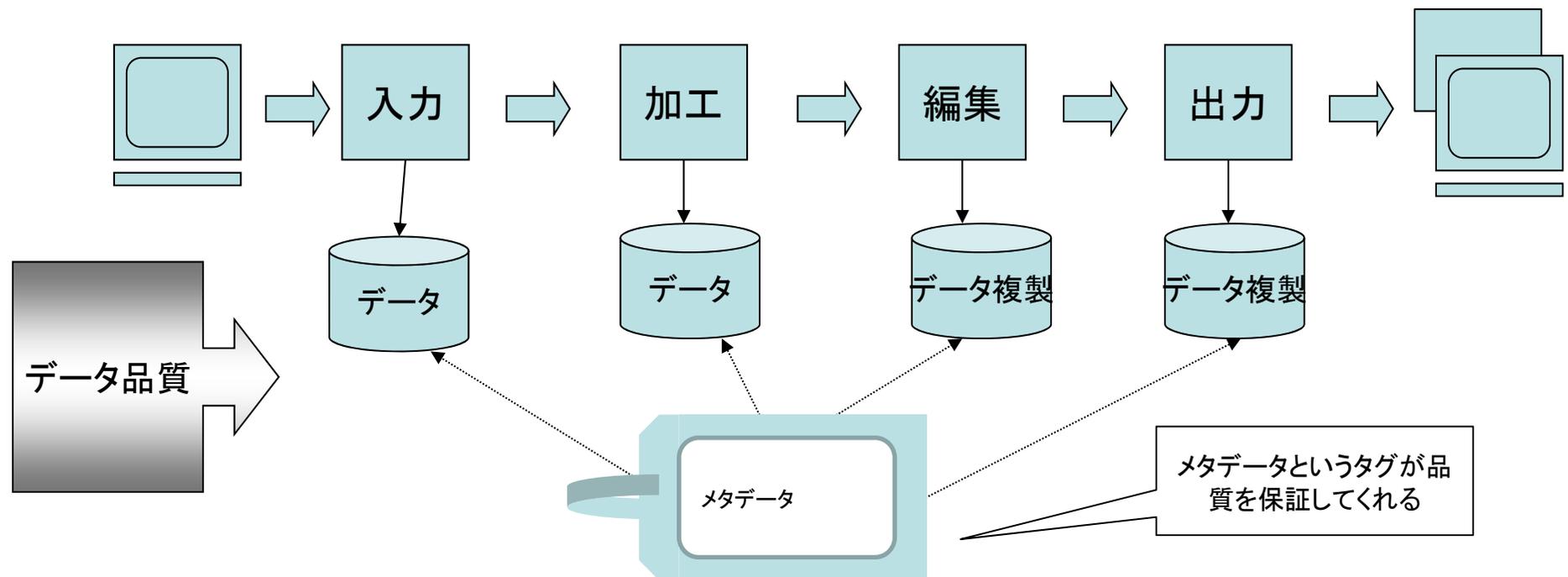


質の高いデータとは

- 正確である
- 完全である
- 現在のデータである
- 整合性がある
- コンテキストから切り離されていない
- 監視できる

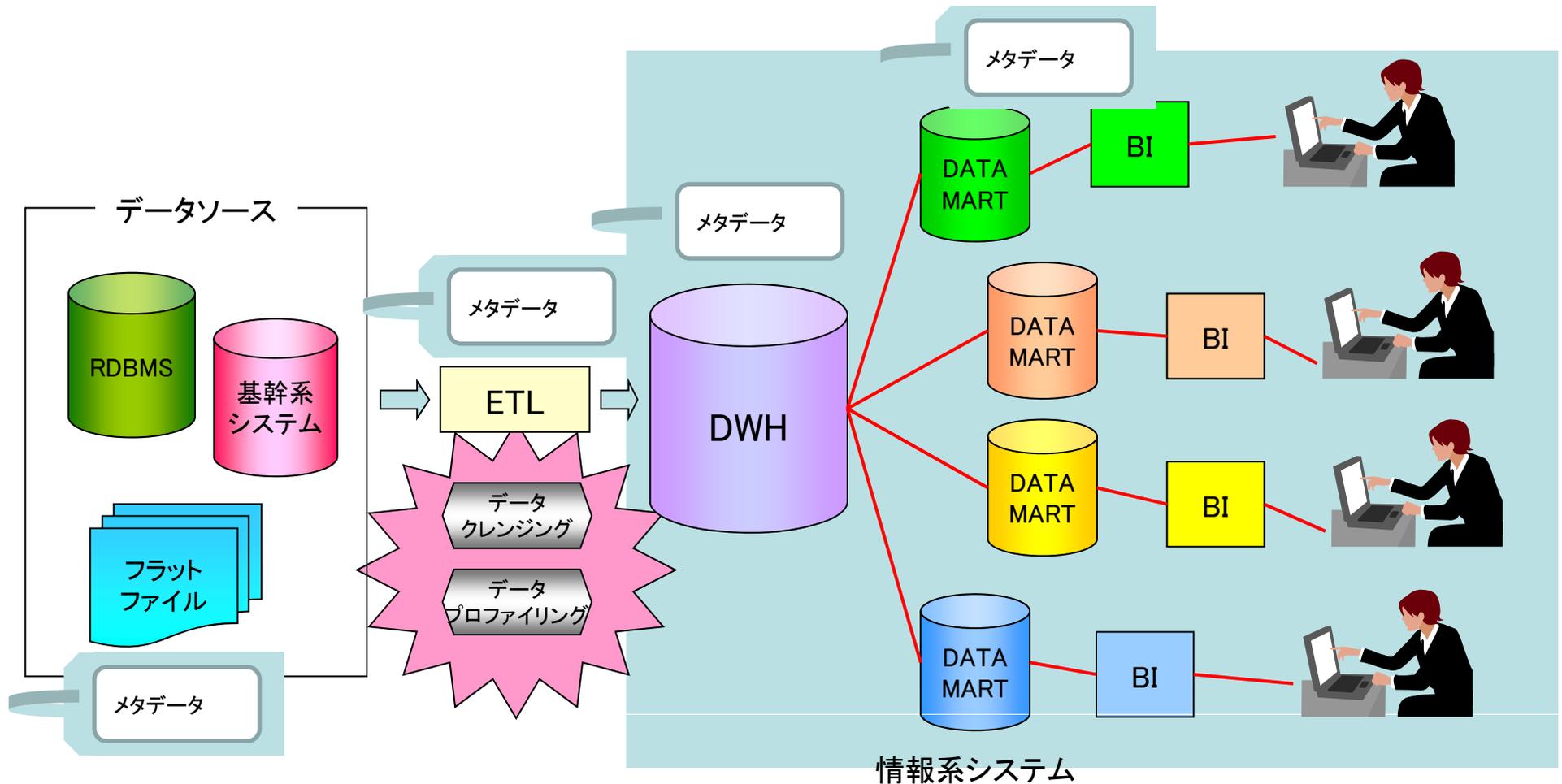
データ品質

- データ品質を形成する要素
 - ◆ メタデータ品質: データに関する定義、バリデーション、デリベーション
 - ◆ コンテンツ品質: データ定義が正しくても誤ったデータは混入する
 - 誤データ投入 (システムでチェックできない)
 - One fact in any place



データ品質

- 基幹系システムに蓄積されたデータを洗浄してデータの精度を高めてDWHを構築する
- 基幹系システムのデータは汚れている?
 - ◆ システム毎に別マスタを作成していて同期が取れていない
 - ◆ 長年運用している間に、コード体系が変わったりして、過去の実績データ間でのアンマッチが発生している

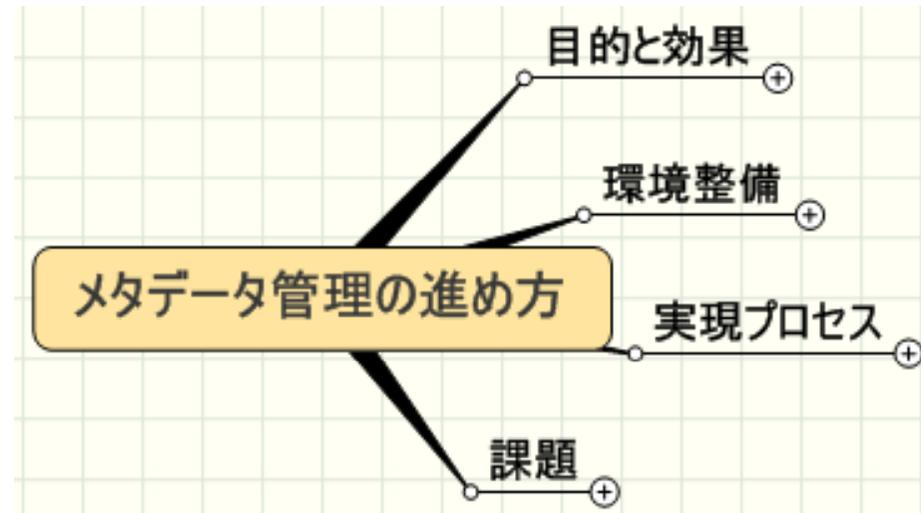


メタデータ整備の手順

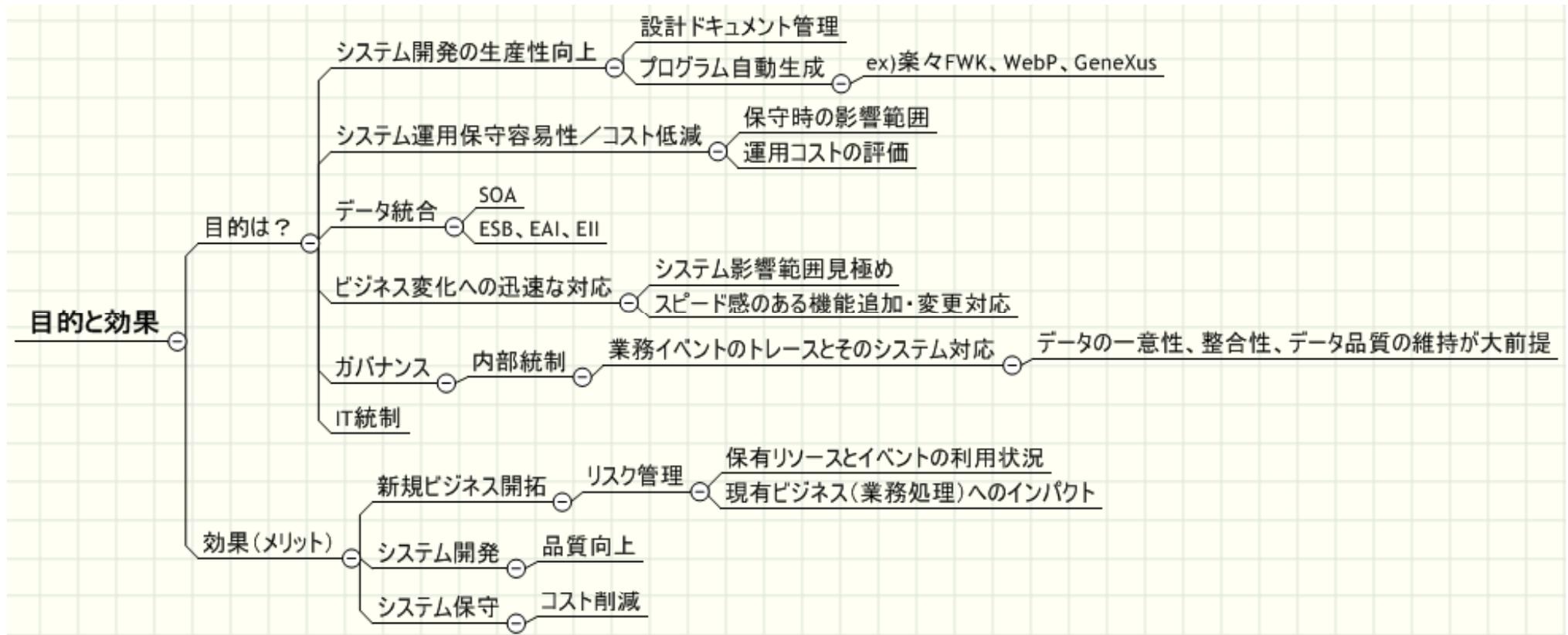
1. 基礎データ(オペレーショナル・データ)整備
 1. 目的・効果
 2. 環境整備
 3. 実現プロセス
 4. 課題
2. ナレッジ共有の仕組み

基礎データ(オペレーショナル・データ)整備の手順

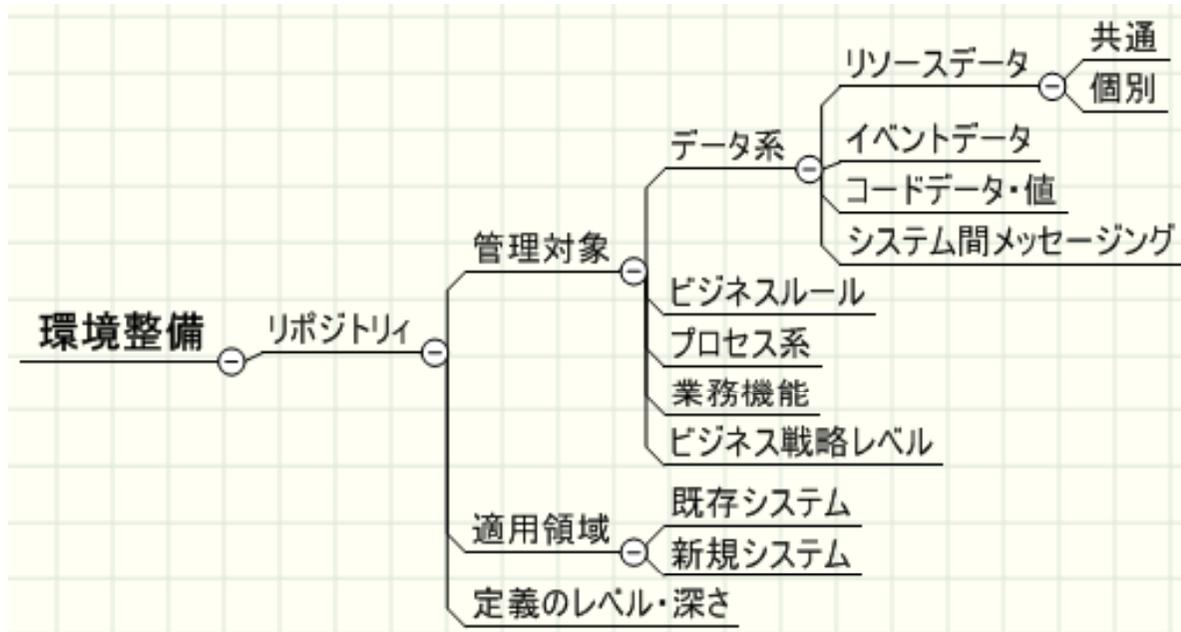
- 目的と効果
- 環境整備
- 実現プロセス
- 課題



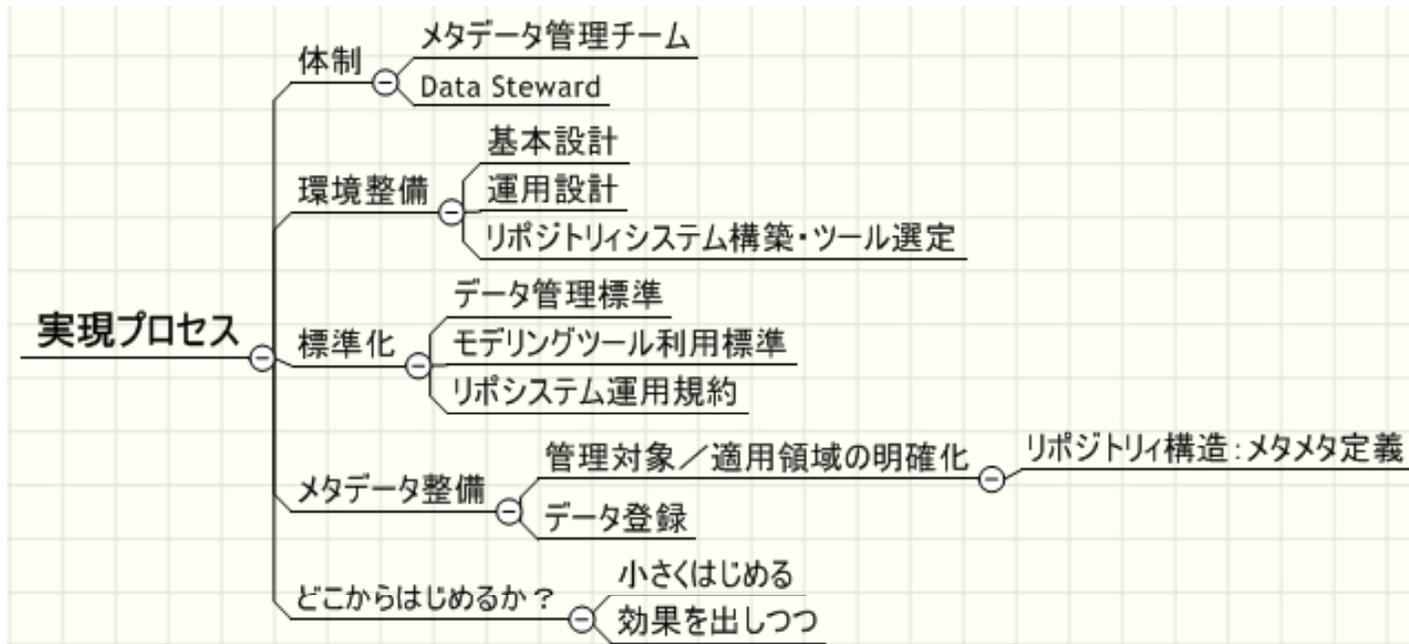
目的と効果



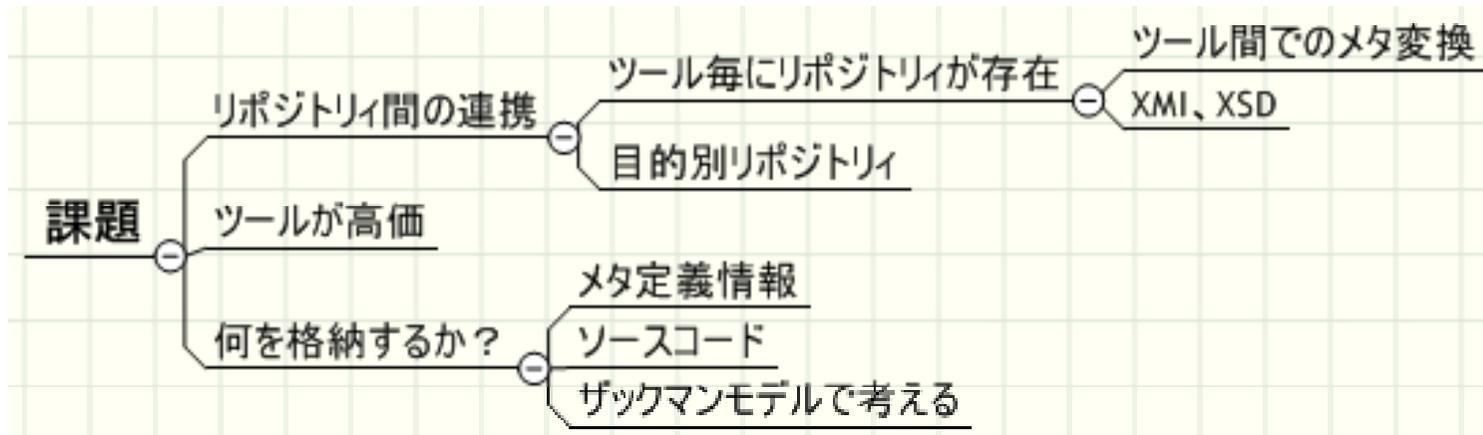
環境整備



実現プロセス



課題

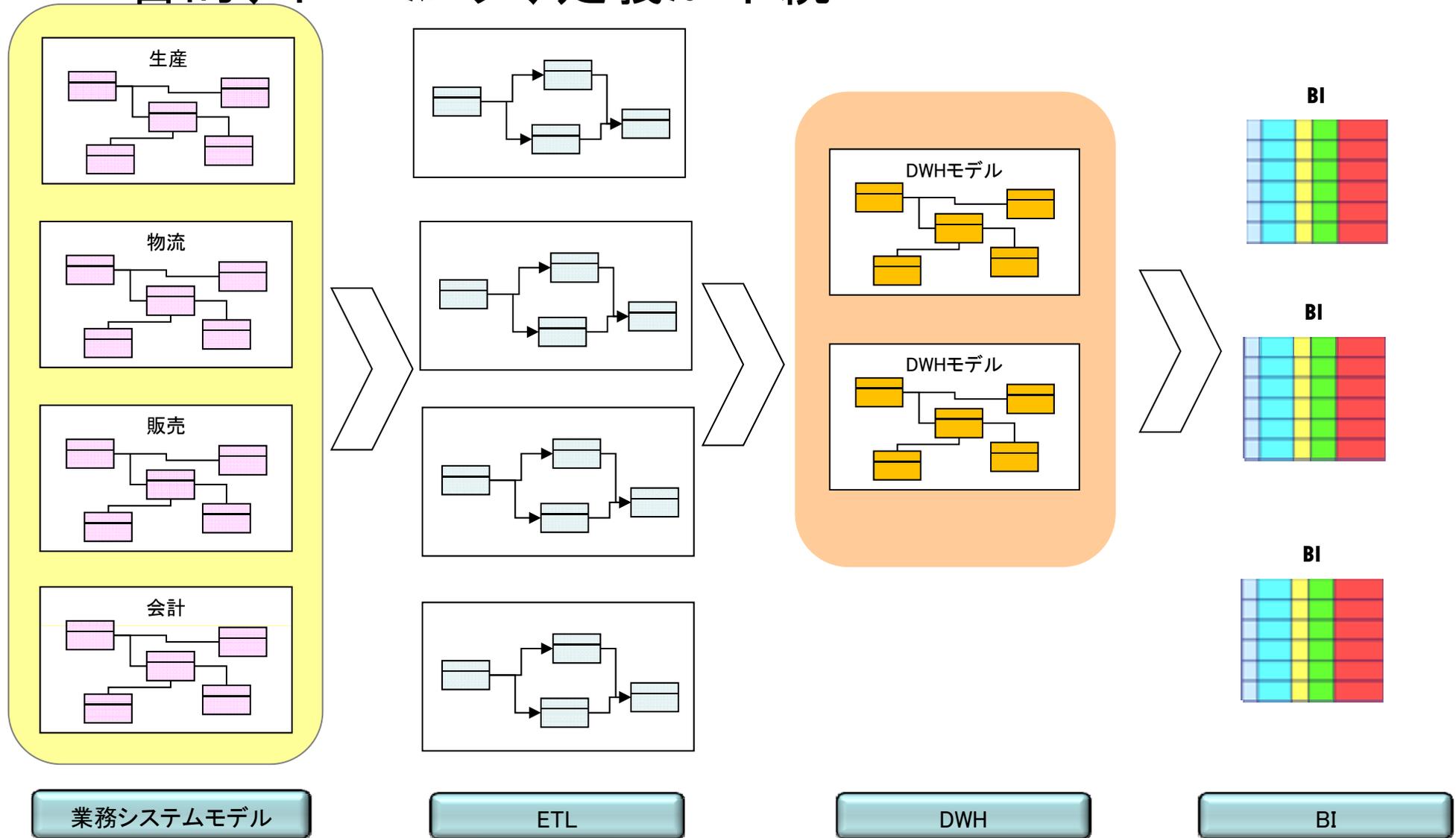


メタデータ統合化アプローチ

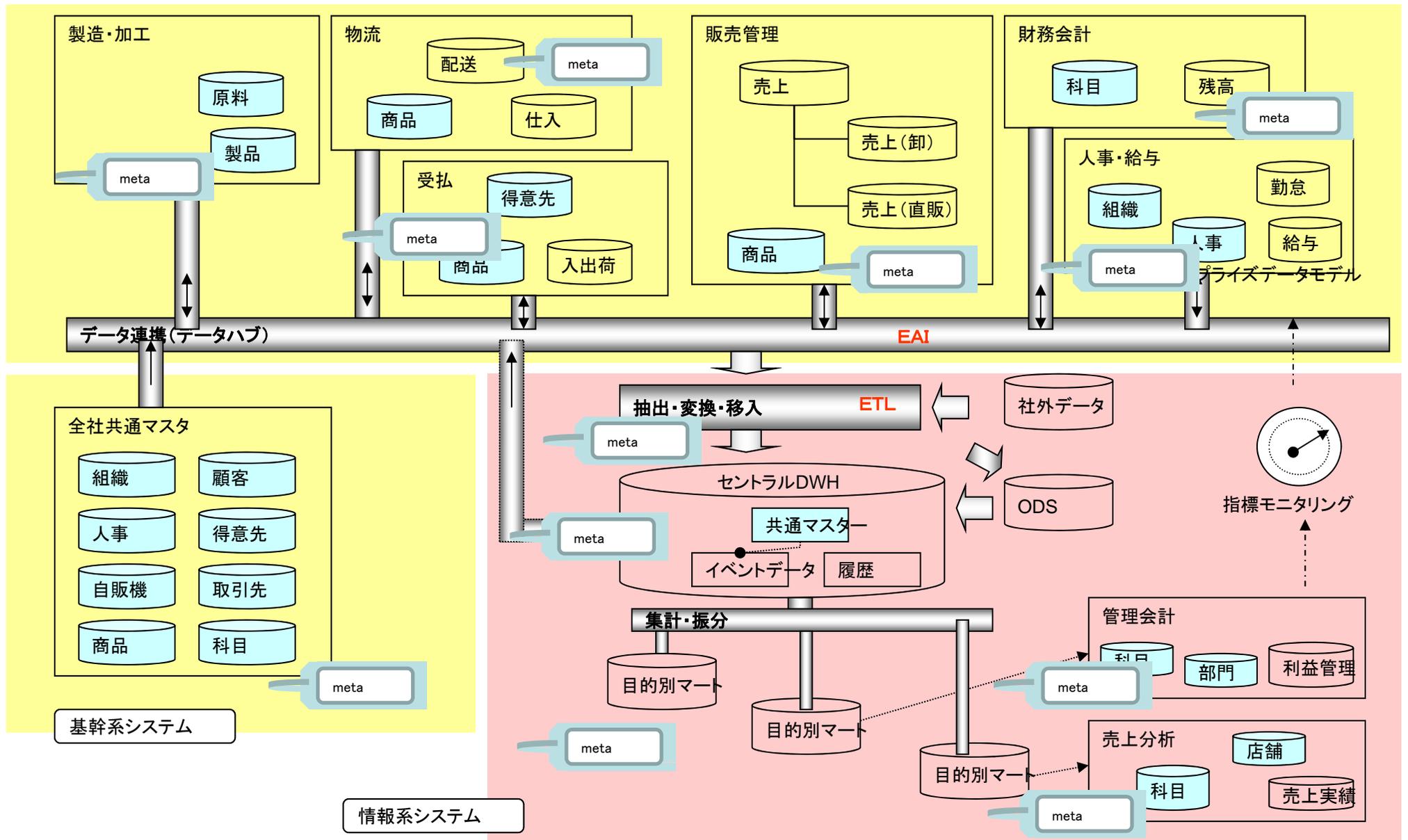
1. サイロ状態
2. 部門別管理から超企業レベルの管理へ
3. メタデータの分類・層別
4. 統合手法
5. メタデータ分類毎の統合とメタデータ分類間の連携
6. メタデータガバナンス効果

サイロ状態のメタデータ

■ 目的、ネーミング、定義が不統一



データアーキテクチャー（データの配置）とメタデータ



メタデータの分類・層別

- 発生源ベースでの基本定義
 - データ発生源での意味定義:ドメイン定義
- 発生源から最終消費までの経路・加工条件
 - 変遷、加工:データリネージ
- システムと利用者視点でのモデル相関関係
 - 横の繋がり、リレーションシップ:データモデル

基本(モデル)定義

- メタデータの定義はここから始まる

■エンティティ

- 名称
- 目的
- 所有者(部門)

■データ項目

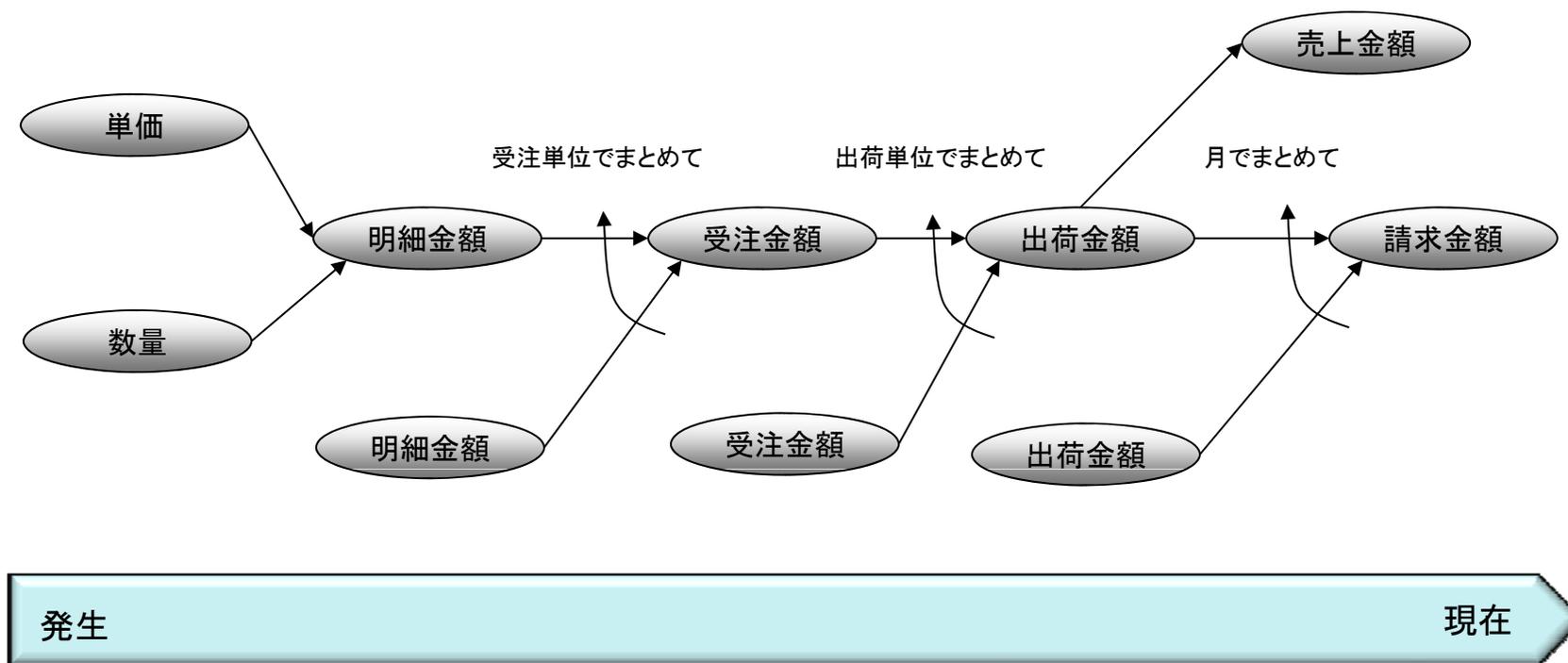
- 正式名称
- 別名
- 発生部門
- 本籍
- 意味定義
- 鮮度
- 加工式
- 所属エンティティ

データリネージ(系譜)

- データを有効活用するために、データがどのように抽出、変換されているかを追跡し、それを記録していくことは、不可欠。
- データの抽出、変換を追跡していないと、異常なデータの原因を調べたりデータ統制のために、膨大な追加作業が発生する。
- ETLツール

データリネージ

- そのデータは発生時点からどのような変遷を経て今に至ったか



ETLツール例示 Informatica PowerCenter



- GUIで情報に必要なデータを遡れる
- ソースの変更に応じて正確で最新のメタモデルが変更される
- ユーザードメインと関連のあるビューを定義できる

Design Time Lineage for: DM_MM_CUSTOMERS
This report generated: Monday, May 31, 2004 7:39 PM
Limit lineage to: All domains Mersche Motors Busin... Upstream and downstream

Design Time Lineage for: DM_MM_CUSTOMERS
This report generated: Monday, May 31, 2004 8:03 PM
Limit lineage to: All domains Mersche Motors Oracl... Upstream and downstream

Design Time Lineage for: CUST_NAME
This report generated: Monday, May 31, 2004 8:07 PM
Limit lineage to: All domains All repositories Downstream

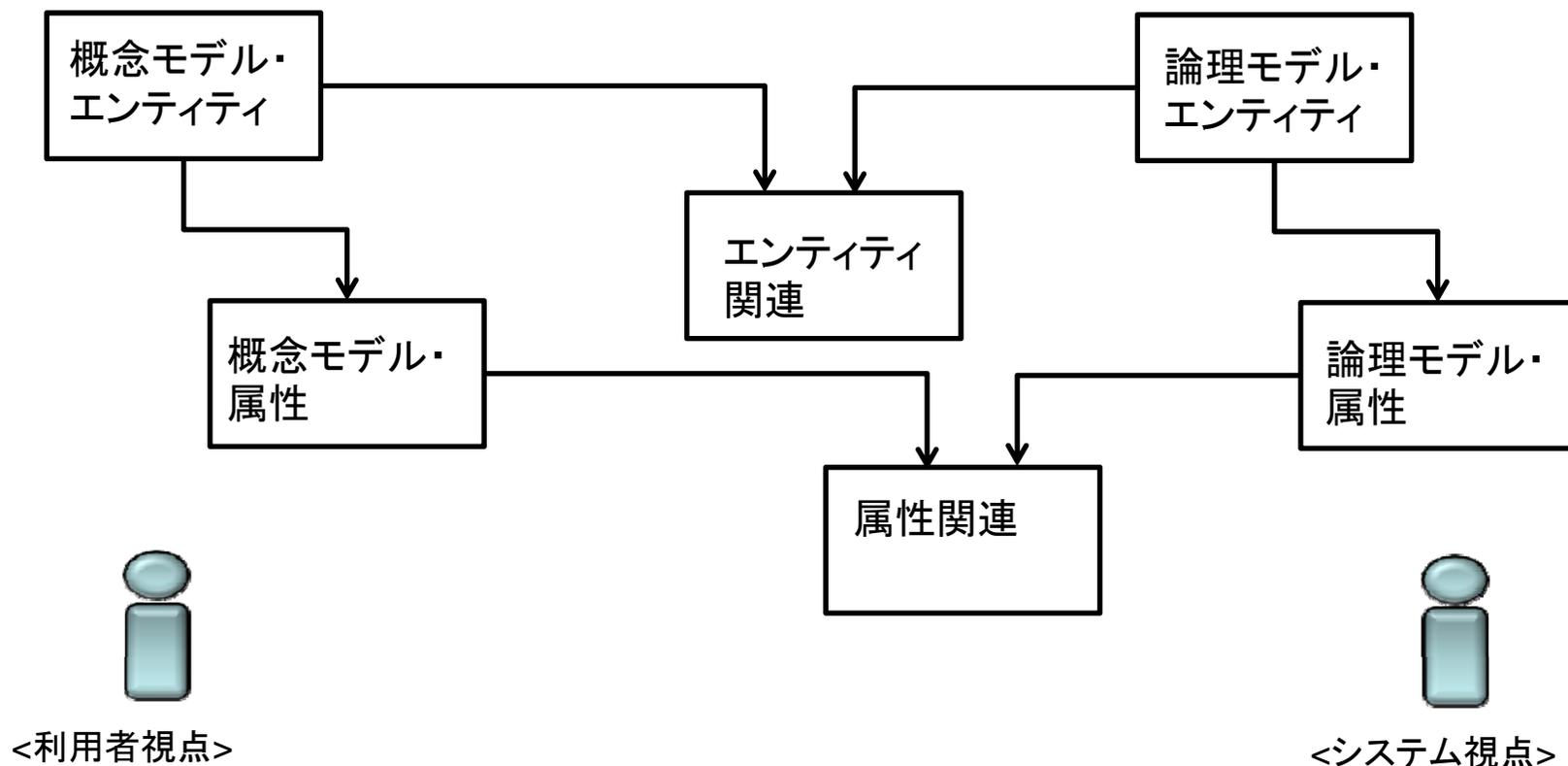
Field Details			
Field Level Lineage for: MM_CUSTOMERS			
Participating Fields:			
CUST_NAME	Parent:	Field Name:	Repository:
	MM_CUSTOMERS1	CUST_NAME	Mersche Motors Powercenter
	Output Fields: This section lists the downstream fields connected to the field selected at left.		
	V_CUSTOMERS	CUST_NAME	Mersche Motors Oracle database
	MM_CUSTOMERS	CUST_NAME	Mersche Motors Powercenter
	MM_CUSTOMERS	CUST_NAME	Mersche Motors PowerAnalyzer

モデル相関関係

■ 利用者視点とシステム視点を関連付ける

◆ 概念モデルー論理モデル

◆ BI用ビューー論理モデル



メタ統合手法

- 共通定義、ネーミング標準
- ツール間のブリッジを構築
- 組織単位から企業レベルのための連邦型リポジトリ
- 複数の組織からのアクセスを可能とする1リポジトリ

MIR(Meta Integration Repository)



開発用メタ

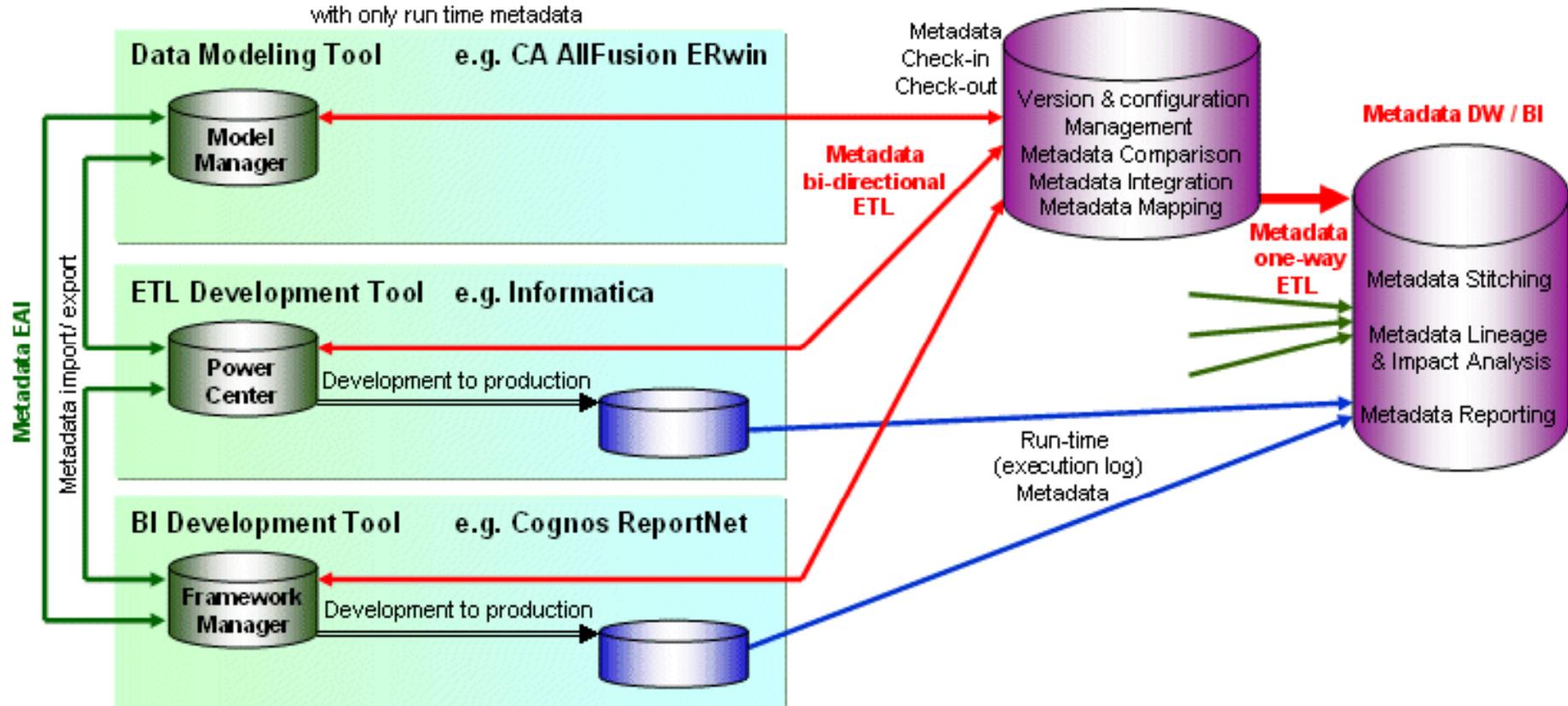
オペレーショナルメタ

ライフサイクルメタ

分析用メタ

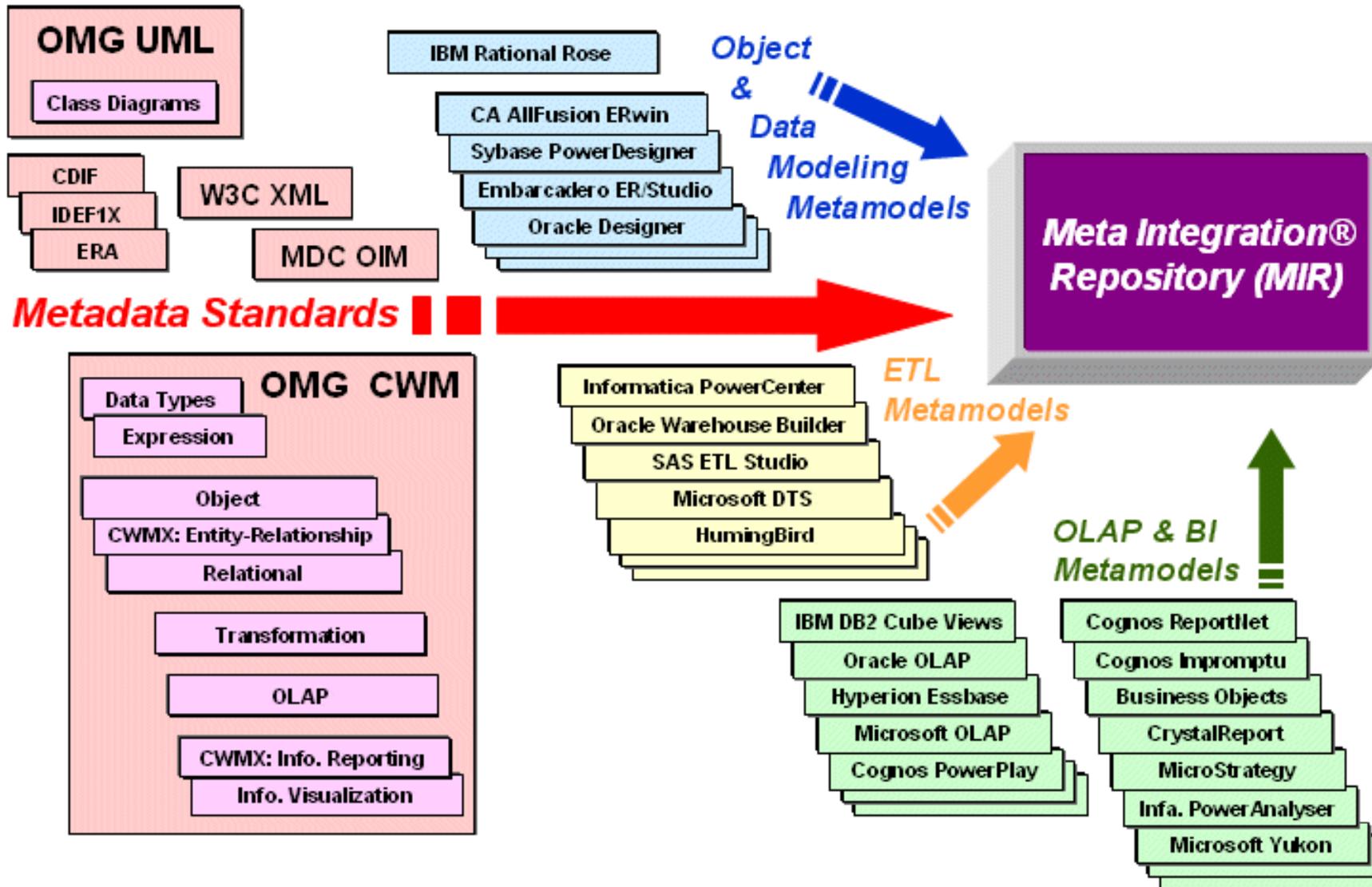
The development and operational metadata repositories can be the same product (development vs. production instance) or the operational repository can be a specific product with only run time metadata

The life cycle and analysis metadata repositories can be the same product.



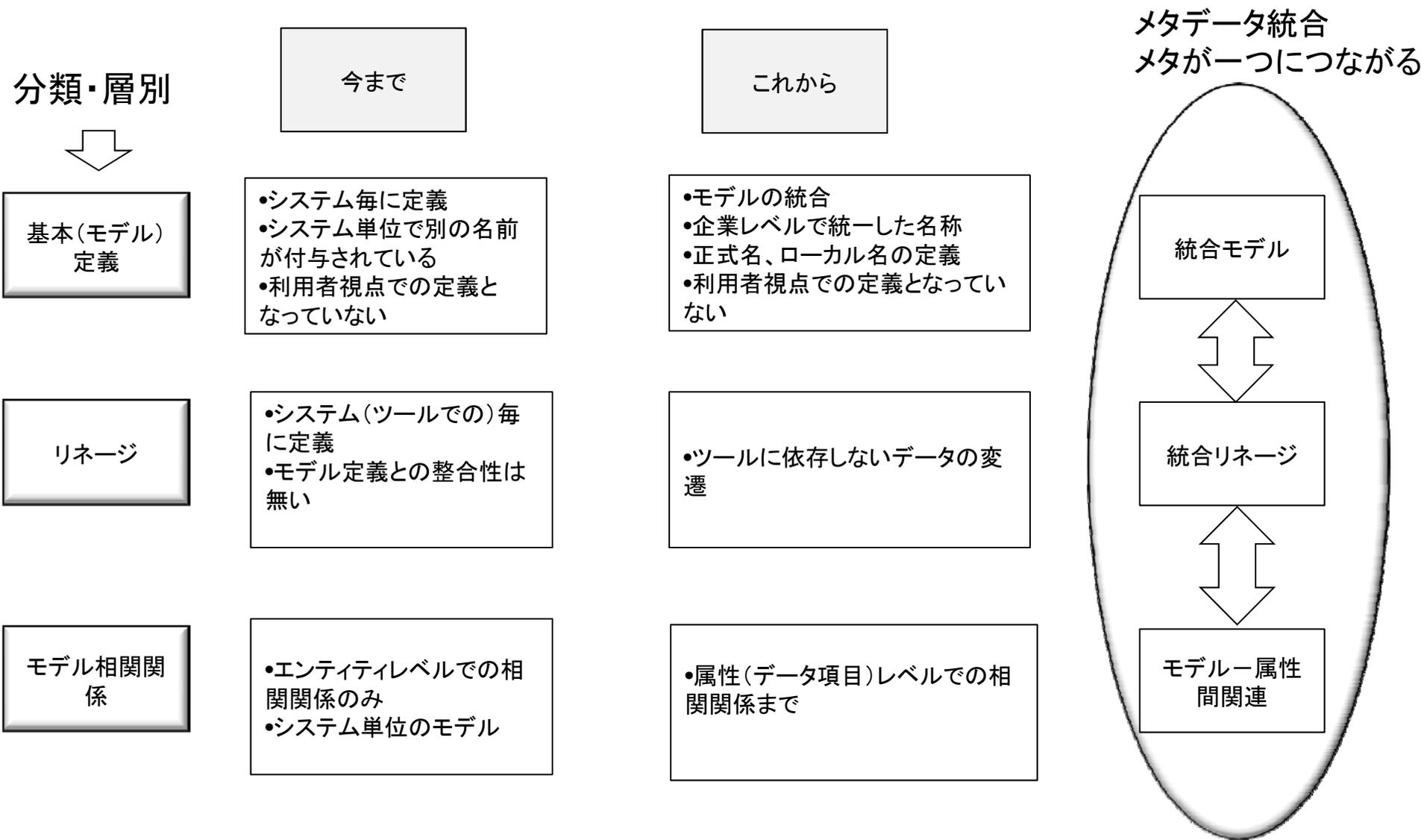
MI社HPより

MIR



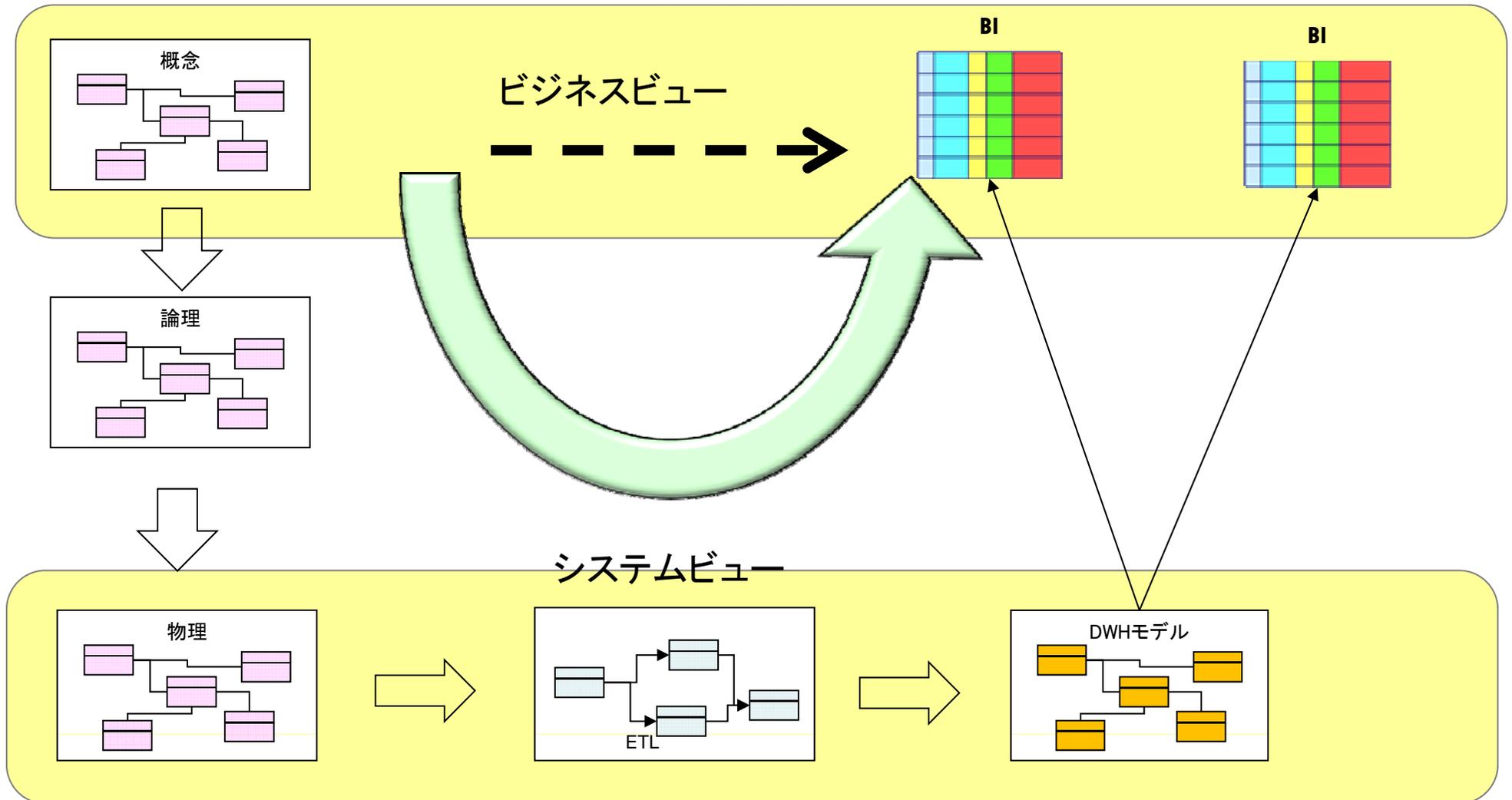
MI社HPより

メタデータ分類毎の統合とメタデータ分類間の連携



視点別メタデータの連携

- 概念モデルからBIまでメタデータが連携する
- ビジネスビューとシステムビューが連携する



メタ統合時のポイント

- 共通定義

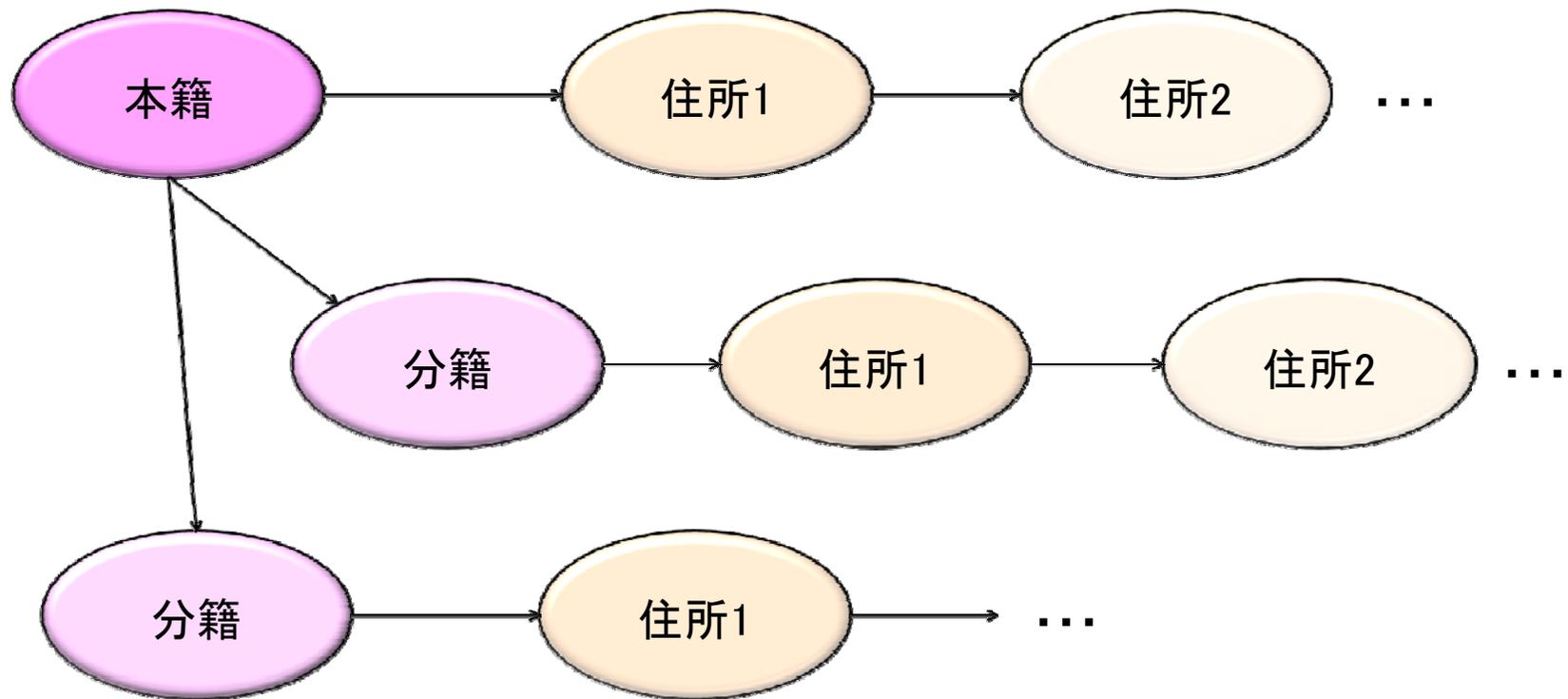
 - ◆ マスタ・ハブデータ

 - ◆ トランザクション・ハブデータ

- ドメイン、本籍管理

- 別名、ロール名

ドメイン管理



メタデータガバナンス効果

- データ項目追加・変更時の影響分析に要する時間短縮
- データの意味把握ミスによる開発時の手戻り率減少
- 開発・保守ベンダーのコスト見積り妥当性判断向上
- ユーザからのシステム問い合わせ応答時間短縮
- システムでのインタフェース不正トラブル減少
- ビジネス要件変更時のインパクト調査時間短縮
- データ品質測定指数向上
- データ収集コスト減少
- 関係者間でのコミュニケーション・コスト減少
- データ検索時間短縮

まとめと提言

- メタデータは、システム単位、ツール毎に乱立状態だ
- メタデータは、基本、リネージ(系譜)、モデル相関に分類・層別される
- ビジネスビュー→システムビュー→ビジネスビューの循環を確立することが必要だ
- 非構造化データを付加してナレッジ共有へと進化させる
- メタデータはデータの品質保証書である
- メタデータガバナンスは、一朝一夕には為し得ない(ボディブローの如く効いてくる)

謝辞

ご静聴有難うございました。



講演内容に関してご質問がございましたら、下記までお問い合わせください。

<http://dataarch.co.jp>

[E-mail: mano@dataarch.co.jp](mailto:mano@dataarch.co.jp)