

ホワイトペーパー

エンタープライズ メタデータ リポジトリの構築

CDO Technologies 社 Ron Lewis

2009 年 12 月

エンバカデロ・テクノロジーズ

〒102-0072 東京都千代田区飯田橋 4-7-1 ロックビレイビル 8F
TEL 03-4577-4520 FAX 03-6843-0961

目次

はじめに.....	1
すべてはデータに帰着する.....	1
メタデータによってコンテキストを明らかにする.....	1
メタデータリポジトリは卵をすべて1つのかごに格納する.....	1
第1段階：収集対象を決定する.....	3
リバースエンジニアリングによって主要なレポート要件からデータ収集ニーズを抽出する.....	3
第2段階：メタデータを収集する.....	3
従来の手法.....	3
主な課題.....	3
解決策.....	4
改善策：ツールを利用する.....	4
第3段階：メタデータリポジトリに格納する.....	5
適切なツールを選択する.....	5
執筆者が選択したツール.....	6
メタデータ収集ツール.....	6
データベースポーリングツール.....	6
アプリケーションプロファイル作成ツール.....	6
まとめ.....	7

はじめに

この数年間、数多くの Web セミナー、テレビ会議やホワイト ペーパーで、メタデータを管理することの重要性が議論されてきました。また、"メタデータ リポジトリの構築" にも注目が集まっています。"流行のメタデータ" についてのホワイト ペーパーの 1 つで終わらないよう、また、特定のメタデータ リポジトリ ソリューションの宣伝にならないよう、このホワイト ペーパーでは、メタデータおよびメタデータ リポジトリの概念の基礎的な定義を行い、メタデータの収集とエンタープライズメタデータ リポジトリへの格納について簡単な方法論を紹介します。

すべてはデータに帰着する

最初に、データと情報の違いを定義したいと思います。データとは、手を加えない、生の事実ではありません。その "生の事実" を意味のあるコンテキストに置いて操作または処理することで、**データは情報**に変化します。データはよく組織の "宝" であると言われる。これは正しいたとえです。組織のデータはその組織にとって、一般的に非常に貴重なものだからです。財政状態が健全か、利益が上がっているかは、多くの場合、企業体がデータ リソースをうまく活用しているかどうかと関連しています。

技術の進化に伴い、経営幹部たちは、組織の効率にとってデータが重要であることを強く認識するようになってきました。企業の効率を上げる方法をより深く理解するために、方法論やフレームワーク（リーン エンジニアリング原則やザックマン フレームワークなどがよい例です）が重要視されています。また、ビジネス プロセスとそのビジネス プロセスに必要な企業のデータとのマッピングも注目されています。その結果として、メタデータ管理の必要性が浮かび上がってきました。

メタデータによってコンテキストを明らかにする

メタデータとは、データについて記述したものであり、データをより効率的に使用するためのものです。多くの場合、メタデータは、ビジネス メタデータと技術メタデータのどちらかに分類されます。ビジネス メタデータでは、分類法について記述し、ビジネス ルールを明示し、共通の語彙を確立します。それにより、データをビジネス上のコンテキストに結び付けることができます。逆に、技術メタデータでは、データの情報源、属性、ドメイン、命名、動き、利用規則について記述します。要するに、データがどのようなコンテキストで意味を持つかを明らかにするためにメタデータが使われるということです。

メタデータ リポジトリは卵をすべて 1 つのかごに格納する

メタデータに関して言うと、ことわざとは逆に、卵をすべて 1 つのかごに入れる方が有効です。メタデータが適切に集められている場合は特に、データを活用して効率や収益性を大きく向上する新しい方法がメタデータによって見えてくる可能性があります。収集したメタデータから最大限の価値を得るには、索引を付けたり分類したり検索したりしやすい方法でメタデータを格納することが重要になります。メタデータ リポジトリはそれを助けるものです。メタデータ リポジトリとは、企業のメタデータの集約、索引付け、分類、保護、アクセスを行うためのシステムです。

これは簡単な仕事ではありません。メタデータにはさまざまな種類があり、メタデータを表現する方法も数多くあるため、メタデータを一元的に格納するのは困難です。先述のように、メタデータには

基本的にビジネス メタデータと技術メタデータの 2 種類があります。下に、重要なメタデータと、それを表現するために通常使われる形式を、例としていくつか挙げます。もちろん、すべてを網羅したリストにはなっていませんが、メタデータ収集作業について知るのに役立つでしょう。

重要なビジネス メタデータ：

- ビジネス語彙。組織で一般的に使われている用語を記述したものです。ビジネス定義から構成されます。
- ビジネス定義。一般的な用語の一般的な意味を明らかにします。
- ビジネス プロセス。ビジネス目的を軸に行われ、ビジネス ルールによって規定される、ビジネス作業の集合です。
- ビジネス ルール。組織と、組織がビジネス目的をどう達成するかを定義します。

重要な技術メタデータ：

- データ モデル。
- システム カタログ。
- ETL（抽出、変換、ロード）スクリプト。
- データ系譜。
- データ利用規則。

メタデータを表現するための代表的な方法：

- ユースケース。ビジネス プロセスをソフトウェア開発の用語で記述するためによく使われます。
- ビジネス モデル。ビジネス分析者とシステム分析者の間のコミュニケーションを向上するために使われます。
- データ モデル。データ要素間の関係を記述するために使われます。

メタデータ リポジトリ ソリューションは、上記のデータすべてを、検索しやすく、かつ保護された形式で収集できるものでなければなりません。メタデータ リポジトリのセキュリティに関して大ざっぱに言うと、メタデータの価値はメタデータ リポジトリの内容物について認識されている品質と信頼性に比例します。メタデータは非常に貴重なものであり、十分に保護しなければなりません。

第 1 段階: 収集対象を決定する

当然ながら、収集してリポジトリに格納できるメタデータは大量に存在します。何を収集する必要があるかの判断は、間違いなく非常に困難です。収集対象の決定は、エンタープライズ エンジニアリングの作業であると考えられています。エンタープライズ アーキテクトは通常、ザックマン フレームワークなどの形式的な構造化フレームワークを使って組織を記述し、企業についての "why、how、what、who、where、when" を定義します。これは費用も時間もかかる作業ですが、それだけの価値はあります。

リバース エンジニアリングによって主要なレポート要件からデータ収集ニーズを抽出する

多様で構造化が進んだ組織の場合には、企業で使用しているレポートをリバース エンジニアリングし、企業の特徴を示すビジネス プロセスに関連するデータを反復的に収集することから始めるのが簡単です。たとえば医療業務であれば、医者、患者、往診、さらには治療など、いくつかの大きな構成部品に分解することができます。医療業務では、たとえば請求書、治療記録、カルテ請求、HIPAA 解除フォームなど（これらは単に例として挙げただけです）、さまざまな種類の重要なレポートが必要です。まず、最も頻繁に使われる "レポート" を取り上げ、大きな構成部品に関連する仕事を完了するために必要な情報を洗い出します。レポートには、重要なビジネス ルールや主要なビジネス属性がまとめられています。セキュリティ業界における経験から言うと、レポートではビジネス データが最も意味のあるコンテキストに置かれているため、レポートはサイバー攻撃の主要な標的となります。セキュリティについてのこの知識を利用したのが、メタデータのニーズを判断するためのこの方法です。レポートには、ビジネス ニーズとそれに必要なデータとのマッピングが、理解しやすく利用しやすい形で表されているため、"レポートにはどのような情報が記述されているか" および "そのレポートを作成するには何が必要か" という 2 つの問いに対する答えが、収集すべき最も重要なメタデータを明らかにするための出発点となります。

第 2 段階: メタデータを収集する

従来の手法

多くのエンタープライズ アーキテクトは、まずビジネス プロセス モデルを作成し、それからそのビジネス プロセス モデルを概念データ モデルにマッピングします。この方法では、ビジネス アーキテクトは、主要なビジネス分析者へのヒアリングを行い、ビジネス モデルを引き出して確認する必要があります。また、データ アーキテクトは、主要な技術担当者へのヒアリングを行い、概念データ モデルを作成して確認する必要があります。モデルの作成とマッピングが終わった後、エンタープライズ アーキテクトは、システム間の依存を特定しながら、それをシステムやデータ リポジトリにマッピングします。現状について定義した、相互に関連付けられたこのメタデータは、収集、分類、索引付けされて、メタデータ リポジトリに格納されます。

主な課題

コストの管理。既存環境 ("現状" など) について手作業で記述するという最初のメタデータ収集作業は、費用面でも時間面でもコストがかかります。自動化できないヒアリングの工程は混乱の元です。

ヒアリング工程には、生産低下というコストも隠れています。収益を生み出す仕事から重要な人員が引き離されるからです。進化が早く要求の厳しい環境では、この損失は壊滅的な影響を与えかねません。

対象範囲の管理。最初にデータ収集作業をするときには、範囲に関して一般的に 2 つの傾向が見られます。すべてを収集しようとするか、詳細部分にこだわりすぎるかです。いずれの場合も勢いがそがれ、コストが跳ね上がることとなります。

データ収集から脱線しないこと。明らかに非効率な問題が見つかった場合などには特に、データ収集作業を中断して矛盾点の対処や修正をしたくなりがちです。しかし、その変更により組織全体に波及効果が起こり、多大な悪影響が生じかねません。

解決策

これらの課題に対処するためにほとんどの状況で役立つような解決策をまとめると、次のようになります。

- 重要なことに注力する。組織の一番の目的から開始し、その後、組織の収益における割合に応じて作業量を定めます。
- 最大限に自動化を進める。手作業で行うヒアリングをできるだけ少なくし、レポートの分析やアプリケーションのリバース エンジニアリングや最新のソフトウェア開発文書などから、できるだけ多くのビジネス情報を収集します。注意：ソフトウェア開発成果物のすべてが、実稼働環境に配置されたものと一致しているとは限りません。
- データ収集時にはデータ リエンジニアリングやプロセス改善を行わない。データ収集の作業だけを行ってください。すべてのメタデータの収集、分類、索引付けが終わった後で、企業全体にまたがる効率の悪さがビジネス分析者に明確に見えてくるのがよくあります。

改善策: ツールを利用する

データ収集作業の総合的な効率を向上しながらも時間と費用を削減する第一の方法は、技術を利用することです。レポートから出発して、それを導き出すプロセスにさかのぼるのが最善です。レポートとそれに関連して収集したメタデータは、プロセスの妥当性を確認する優れた手段であり、ヒアリング工程に参加する全員にとっての共通の知識の枠組みとなり、企業全体を完全に把握するために必要なヒアリングの数を最小限に抑えるのに役立ちます。

ツールを使って次のことを行います。

- 社内のデータベース サーバーをできるだけ多く洗い出し、各データベース サーバーに含まれる物理スキーマをリバース エンジニアリングします。
- エンタープライズ データ モデルを 1 つ作成し、リバース エンジニアリングで抽出した物理スキーマそれぞれを、そのマスタ モデルに追加します。
- スキーマ メタデータをスプレッドシートにエクスポートし、各エンティティについてわかっている事柄と関連する属性とを適切なデータ アーキテクトに書き込んでもらいます。
- アプリケーション プロトコル モニタを適用して各アプリケーションのプロファイルを取り、プログラム ロジックとして記録されたビジネス ルールを捕捉します。つまり、データ利用規則を捕捉することになります。
- アプリケーション ソース コードからビジネス ロジックを抽出します。

- ETL（抽出、変換、ロード）スクリプトのプロファイルを取って、変換ルールとデータ系譜メタデータを捕捉します。これは、各データベース サーバーにデータベース プロファイリング ツールを適用して使用状況を捕捉することで行います。
- 既存のビジネス プロセス モデル（BPM）を標準の BPM ツールにインポートします。
- データ使用状況のプロファイルを取ります。これは、長い間使われていないデータや利用頻度の高いデータ（データベース、テーブル、要素）の特定に重点を置いて行います。
- 捕捉したメタデータが一貫しているかを評価します。矛盾があれば特定し、注釈を付けます。

第 3 段階: メタデータ リポジトリに格納する

ここまでの作業が終わった時点で、相当な量のメタデータが存在するはずですが、メタデータから価値を取り出すには、共通のリポジトリにインポートし、分類し、索引を付けて、適切なビジネス分析者、システム分析者、データ分析者が使える状態にしなければなりません。

メタデータ リポジトリは、構築済みのものがいくつも提供されています。メタデータ リポジトリ ソリューションには、非常に重要な次の機能を持つものを選択しなければなりません。

- 非構造化コンテンツを管理できる。
- 適切なアクセス制御と監査機能を持っている。
- さまざまな形式のものをインポートでき、XML のような標準形式でエクスポートできる。
- 各メタデータ資産について説明を書いて関連付けることができる。
- 適切なメタデータを検索し評価するための直観的な方法を備えている。

メタデータ リポジトリに格納する手順はソリューションによって異なりますが、方法論はよく似ています。各ソリューションは、メタデータ収集ツール間でデータをやり取りするためのインポート/エクスポートのメカニズムを備えているはずですが、もともとリポジトリ ソリューションでサポートされているツール以外の収集ツールでは、XML などの一般に知られた形式を中間媒介物として使用することができます。たとえば、データの命名について記述したメタデータを Microsoft (MS) Excel などの厳密に構造化されていないツールに記録している場合には、メタデータをコンマ区切りの CSV ファイルにエクスポートし、それをサポートするデータ モデリング ツールにインポートすることができます。ほとんどのデータ モデリング ツールでは、メタデータを XML 形式でエクスポートすることができます。MS Excel ファイルは通常、リポジトリ内で "非構造化コンテンツ" として扱われますが、適切なメタデータをデータ モデリング ツールにインポートし、スキーマ メタデータとしてエクスポートすることで、構造化コンテンツに変換することができます。この例ではもちろん、スプレッドシートに表現されたメタデータが論理データ モデルまたは物理データ モデルに関連付けられていると想定しています。

適切なツールを選択する

メタデータ収集に適切なツールを選択することは非常に重要です。直観的で使いやすいツールでなければなりません。さらに重要なのは、ツール同士がしっかりと統合されて、ツール間でデータをやり取りするための作業が最小限になっていなければならないことです。（メモ：ツールからデータをエクスポートして別のツールへインポートするたびに、メタデータが失われたり変化してしまう危険性があります。）また、業界標準として受け入れられているツールを選ぶ必要があります。そうすれば、

どのメタデータ リポジトリ ソリューションを選んだとしても、そのツールがサポートされているはずだからです。

執筆者が選択したツール

使いやすく、信頼性が高く、デモをしやすいツールを選ぶことが重要です。私は Embarcadero のソリューションを本当に気に入っています。新しい Embarcadero All Access ソリューションで統合が行われたことで、その選択が間違っていないと改めて思いました。私はこのツールセットを補完するものとして NitroSecurity の NitroView Application Protocol Monitor (APM) を使用しています。既存アプリケーションのプロファイルを取って .NET や Java のコードとして表現されたビジネス ロジックを抽出するのに必要な、可視性を得るためです。

メタデータ収集ツール

Embarcadero® ER/Studio® Data Architect。物理スキーマのリバース エンジニアリング、関連する概念モデルおよび論理モデルの構築、スキーマに関するメタデータの管理に使用します。企業の変換、利用、データ系譜メタデータの捕捉や管理に適しています。リバース エンジニアリングにより抽出したスキーマはサブモデルとしてマスタ データ モデルに追加でき、マスタからはマスタ データ カタログを自動生成することができます。ER/Studio Data Architect ではメタデータ タグの作成をサポートしていますが、これは、HIPAA や FISMA 向けに PHI や PII をそれぞれ特定するなど、法的な要求事項を満たすために役立ちます。また、エイリアスの作成や管理もサポートしているので、データガバナンスの基盤としても利用できます。

Embarcadero® ER/Studio® Business Architect。既存ビジネス プロセス モデルのインポートと新しいモデルの構築に使用します。

Embarcadero® ER/Studio® Enterprise。ER/Studio の最新バージョンでは、Data Architect と Business Architect を ER/Studio Repository に統合することができます。その結果、ビジネス プロセス メタデータとデータ メタデータを Web からアクセス可能なリポジトリ内に共存させることが可能になりました。

データベース ポーリング ツール

Embarcadero® DB Optimizer™。セッションを捕捉し、それをデータベースやテーブルでフィルタリングする機能を備えています。データ ウェアハウスの構築に不可欠な、主要パフォーマンス パラメータを捕捉するのに適しています。また、どの ETL スクリプトが使われているかの特定や、データ系譜の妥当性確認を行う機能も提供しています。

アプリケーション プロファイル作成ツール

Embarcadero® JBuilder™。既存の J2EE アプリケーションの評価とビジネス ケースへのマッピングに使用します。

NitroView APM。アプリケーション ロールからビジネス機能へのマッピングと、ほぼリアルタイムでのビジネス ルールの公開に使用します。APM の出力は DB Optimizer の出力と相互に関連付けることができるため、既存アプリケーションで接続プーリングを使用している場合やアプリケーション

ユーザー アカウントを共有している場合でも、どのユーザーがどのデータにアクセスしているかを非常に正確に把握できます。

メタデータのほとんどは ER/Studio Repository に格納され、ER/Studio Enterprise に含まれる ER/Studio Portal から利用することができます。Data Architect と Business Architect を 1 つのリポジトリに統合し、さらにその機能を All Access Toolkit で提供したことで、All Access はメタデータの収集や管理に適したツールとなっています。

まとめ

このホワイト ペーパーの目的は、メタデータおよびメタデータ リポジトリの概念の基礎的な定義を行い、メタデータの収集とエンタープライズ メタデータ リポジトリへの格納について簡単な方法論を紹介することでした。理解してほしい要点は次の 3 つです。

エンタープライズ データ資産から最大限の利益を引き出すには、組織はエンタープライズ メタデータを適切に収集し、管理しなければなりません。

企業全体からメタデータを収集する作業はコストがかかりますが、それだけの価値はあります。適切なツールを使用すると、費用を大幅に削減することができます。また、エンタープライズ ビューを構築することによる目に見える恩恵として、高い投資収益率が得られるはずです。

メタデータ リポジトリは、メタデータ管理を進める上で欠かせないツールです。メタデータ リポジトリの選択は重要であり、リポジトリには、非構造化コンテンツの管理や、一連の主要ツールに対するインポート/エクスポート、XML のような標準形式でのインポート/エクスポートなど、主要な機能が備わっている必要があります。



エンバカデロ・テクノロジーズについて

エンバカデロ・テクノロジーズは、1993年にデータベースツールベンダーとして設立され、2008年にポーランドの開発ツール部門「CodeGear」との合併によって、アプリケーション開発者とデータベース技術者が多様な環境でソフトウェアアプリケーションを設計、構築、実行するためのツールを提供する最大規模の独立系ツールベンダーとなりました。米国企業の総収入ランキング「フォーチュン 100」のうち 90 以上の企業と、世界で 300 万以上のコミュニティが、エンバカデロの Delphi®、C++Builder®、JBuilder®といった CodeGear™製品や ER/Studio®、DBArtisan®、RapidSQL®をはじめとする DatabaseGear™製品を採用し、生産性の向上と革新的なソフトウェア開発を実現しています。エンバカデロ・テクノロジーズは、サンフランシスコに本社を置き、世界各国に支社を展開しています。詳細は、www.embarcadero.com/jp をご覧ください。

Embarcadero、Embarcadero Technologies ロゴならびにすべてのエンバカデロ・テクノロジーズ製品またはサービス名は、Embarcadero Technologies, Inc.の商標または登録商標です。その他の商標はその所有者に帰属します。