Metadata Based Impact and Lineage Analysis Across Heterogeneous Metadata Sources

Presentation at the
THE 9TH ANNUAL Wilshire Meta-Data Conference
AND THE 17TH ANNUAL DAMA International Symposium
by
John R. Friedrich, II, Ph.D.
jrfried@idworld.net
Outline

- Concepts
- Harvesting
- Mapping
- Reporting
- Version and Configuration Management
- Metadata Management Strategy
- Standardization
- Proof of Concept
- Conclusion
Caveat

- This presentation is based upon consulting engagements with Meta Integration Technologies, Inc. (MITI) and facilitated using their tools and methods.
- Meta Integration is the leading "Metadata Component Provider" to major database, data integration, business intelligence, repository, and modeling tool vendors. Meta Integration® OEM components & tools manage and analyze the complete enterprise metadata life cycle across heterogeneous vendors. Solutions range from reusable metadata movement components to repository. Over 50 Model Bridges integrate standards like CWM, and the most popular modeling, ETL/EII, and OLAP/BI tools.
- The Meta Integration components may be found already integrated with its partners’ tools including the Adaptive Repository, Cognos ReportNet, Embarcadero ERStudio, Informatica Power Center and SuperGlue, and more in the future.
- All other trademarks, trade names, service marks, and logos referenced herein belong to their respective companies.
Source Systems

- Data Models
- Process & Code Models

ETL

- Source Data Schemas
- Mappings
- Lifecycle Metadata
- Target Data Schemas

Data Warehouse

- Staging Data Models
- Data Schemas
- Mappings
- Cubes, Joins, etc.

Reporting / Business Intelligence

- Data Sources
- Mappings

Computer Aided Software Engineering

- CASE Design
- CASE Repository
- Forward/Reverse Engineering
- Applications

Extract Transform and Load Development

- ETL Design
- ETL Repository
- ETL Engine

Computer Aided Software Engineering

- CASE Design
- CASE Repository
- Forward/Reverse Engineering

Business Intelligence Design and Reporting

- BI Design
- EAI Repository
- BI Reporting

Target Data Schemas

Process & Code Models
Harvesting

- Metadata survey
  - Source/Target survey
  - Data process survey
  - Reference and standards survey
- Metadata exchange methodology
- Repository Structure
- Extraction
Mapping

- Data process mapping
  - EAI among models
  - ODS to ETL to DW Staging
  - Staging to DW
  - DW to BI
- Metadata management mapping
  - To/From standards and reference models
### Source Message

**Message Class:** Employee  
**Attribute:** Employee Number  
**DataType:** Char(10)  
**Attribute:** Name  
**DataType:** Char(50)  
**Attribute:** Address  
**DataType:** Text()  
**Attribute:** Age  
**DataType:** Number(2)  
**Attribute:** Race  
**DataType:** Char(2)

### Target Schema

**Table:** Person  
- **Attribute:** Unique ID  
  - **DataType:** LongInt  
- **Attribute:** Name  
  - **DataType:** Varchar(50)  
- **Attribute:** DOB  
  - **DataType:** Datetime  
- **Attribute:** Ethnicity  
  - **DataType:** Char

**Table:** Employee  
- **Attribute:** Unique ID  
  - **DataType:** LongInt  
- **Attribute:** Employee Number  
  - **DataType:** Varchar(10)  
- **Attribute:** Date of Employment  
  - **DataType:** Datetime  
- **Attribute:** Status  
  - **DataType:** Char(10)

**Table:** Address  
- **Attribute:** Unique ID  
  - **DataType:** LongInt  
- **Attribute:** Address Line 1  
  - **DataType:** Varchar(30)  
- **Attribute:** Address Line 2  
  - **DataType:** Varchar(30)  
- **Attribute:** Address City  
  - **DataType:** Varchar(30)  
- **Attribute:** Address State  
  - **DataType:** Varchar(2)  
- **Attribute:** Address Zip  
  - **DataType:** Varchar(9)

**Table:** PersonAddress  
- **Attribute:** Person Unique ID  
  - **DataType:** LongInt  
- **Attribute:** Address Unique ID  
  - **DataType:** LongInt  
- **Attribute:** Address Type  
  - **DataType:** Varchar(10)
Reporting

• Lineage analysis
  • Forward (Impact) analysis
    • ODS → DW Staging → DW → BI Design → Reporting
  • Reverse lineage analysis
    • ODS ← DW Staging ← DW ← BI Design ← Reporting
Version and Configuration Management

- Not “one version of the truth”
- Multiple versions of sources/targets
- Multiple mappings among multiple versions
- Multiple configuration of these versions
Metadata Management Strategy
1. Extract Schema from Oracle ERP into MDR
2. Export schema and transforms from Designer
3. Retrieve Model from MDR
4. Edit in ERwin
5. Post in Model Mart
6. Export complete Model from ERwin
7. Import New Version into MDR
8. Migrate Version from Prior ERP Schema Design
9. New ERP Schema Design Completed
10. Export New ERP Design to ERwin
11. Generate ERP Update DDL
12. Export Source / Target models and Transforms from MDR
13. Import into Informatica Designer
14. Store in Informatica Repository
15. Export schema and transforms from Designer
16. Import New Product into MDR
17. Compare Models in MDR
18. Map Models in MDR
19. New ETL Design Completed
20. Extract Schema from Legacy into MDR
21. Import New Product into MDR
22. Export Source/Target Models from ERwin
23. Import New Product into MDR
24. Compare Models in MDR
25. Map Models in MDR
26. Populate Repository with Source/Target Models
27. Customize Source/Target Models
28. Forward-Engineer to Oracle ERP
29. Develop ETL in Informatica
Identify New meta-data

Extract meta-data from source tool

Import into Meta-Data Repository as first version of existing package

Export meta-data from Repository into target tool format

Comparison and mapping of meta-data

Updated meta-data to interact with or impact another meta-data source

Updated meta-data from Standard meta-data package approval

Updated meta-data to be committed to physical instantiation

Editing of meta-data attributes in ERwin

Manage in Model Mart

Integrate meta-data

Integrated meta-data

Extract meta-data from source tool

Import into Meta-Data Repository as new version of existing package

Export meta-data from Repository into target tool format

Manage in Model Mart

Extract meta-data from source tool

Import into Meta-Data Repository as first version of new package

Export meta-data from Repository into target tool format

Integrated meta-data
Standardization

- Mapping to reference models
- Creation of a “canonical form”, i.e., form that may be transformed to any of the other reference forms.
- Standardization process
Proof of Concept

Implementing An Effective Meta-data Management Environment at Worthington Industries, Inc.

Prepared by John R. Friedrich
In conjunction with
Meta Integration® Technologies, Inc. (MITI) and Worthington Industries, Inc.
as part of a
Meta Integration® Based Proof of Concept
POC Introduction

- Worthington Industries, Inc.
  - Worthington Industries is a leading diversified metal processing company with annual sales of approximately $2 billion.
  - Headquartered in Columbus, Ohio, Worthington employs over 8000 people and operates 61 facilities in 10 countries.
Worthington Data Flows
Meta-data is Everywhere
Browse Structure

Meta-Data Repository

based upon
Meta Integration® Model Bridge (MIMB)
MIMB Web Enterprise Edition v4.1.0 05/05/2004

Documentation Supported Tools About Vendors About Meta-Data Standards

Search Repository Browse Repository Version Manager Model Browser

Worthington / ERP / Oracle / Vendor Subset

(Child Product) Vendor Subset
(Version) 20040510
(Model Format) Meta-Data Imported from CA Erwin v 4.4 Export
(DestinationOf Data Movement Format) Data Movement Metadata
### Meta-Data Repository

**WORTHINGTON INDUSTRIES**

*Based upon Meta Integration® Model Bridge (MIMB)*

**MIMB Web Enterprise Edition v4.1.0 05/05/2004**

**Full MDR Search**

#### Search Repository

**Text Search:**
- [x] As possible words in the field
- [ ] As part of the text of the field
- [ ] As the entire text of the field
- [ ] All occurrences (text is ignored)

**Result Level:**
- [ ] Retrieve only matching models
- [x] Show each matching model element

**Metadata Filter:**
- [ ] Select All
- [ ] Deselect All

#### Repository Management

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product</td>
<td></td>
</tr>
<tr>
<td>Version</td>
<td></td>
</tr>
<tr>
<td>Model/Mapping</td>
<td></td>
</tr>
</tbody>
</table>

#### Metadata Origin

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vendor Name</td>
<td></td>
</tr>
<tr>
<td>Tool Name</td>
<td></td>
</tr>
<tr>
<td>Tool Version</td>
<td></td>
</tr>
</tbody>
</table>

#### Model Management Name Description

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td></td>
</tr>
</tbody>
</table>

#### Data/Object Models

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class/Entity/Table</td>
<td></td>
</tr>
<tr>
<td>Attributes/Column</td>
<td></td>
</tr>
<tr>
<td>Operation/Store</td>
<td></td>
</tr>
<tr>
<td>Procedure/Trigger</td>
<td></td>
</tr>
<tr>
<td>Association/Aggregation</td>
<td></td>
</tr>
<tr>
<td>Data Type/Domain</td>
<td></td>
</tr>
<tr>
<td>Other UML/DEFIX Concepts</td>
<td></td>
</tr>
</tbody>
</table>

#### OLAP/BI Models

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimension</td>
<td></td>
</tr>
<tr>
<td>Dimension Attribute</td>
<td></td>
</tr>
<tr>
<td>Measure</td>
<td></td>
</tr>
<tr>
<td>Cube</td>
<td></td>
</tr>
<tr>
<td>Hierarchy/Level</td>
<td></td>
</tr>
<tr>
<td>Join</td>
<td></td>
</tr>
<tr>
<td>Other OLAP/BI Concepts</td>
<td></td>
</tr>
</tbody>
</table>

#### RDBMS Models

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary/Candidate Keys</td>
<td></td>
</tr>
<tr>
<td>Foreign Keys</td>
<td></td>
</tr>
<tr>
<td>Index/Index Member</td>
<td></td>
</tr>
</tbody>
</table>

#### Mapping/Transformation/ETL Name Description

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapping Model</td>
<td></td>
</tr>
<tr>
<td>Transformation</td>
<td></td>
</tr>
</tbody>
</table>

---

*25 May 2005*  
Copyright © Meta Integration® Technologies, Inc.
Search Results: Found 18 instances.

Worthington / Standards / Vendor / 20040510 / Vendor Standards Meta-Data Imported from CA Erwin v. 4 x / Vendor Standards Meta-Data Imported from CA Erwin v. 4 x

This entity describes the vendor that Worthington Industries does business with. Key attributes are Operating name, organization name and duns number. Business name: Vendor Business alias: Customer Business definition Business rules: Business data steward: Accounting/Finance: Dave Cepek Business acronym:

Worthington / Legacy / ASN / 20040510 / Meta-Data Imported from CA Erwin v. 4 x / Meta-Data Imported from CA Erwin v. 4 x

(Attribute) Ship To Duns Number

Worthington / Legacy / ASN / 20040510 / Meta-Data Imported from CA Erwin v. 4 x / Meta-Data Imported from CA Erwin v. 4 x

(Attribute) Ship From Duns Number

Worthington / Legacy / ASN / 20040510 / Meta-Data Imported from CA Erwin v. 4 x / Meta-Data Imported from CA Erwin v. 4 x

(Attribute) Ship To Duns Number

Worthington / Legacy / ASN / 20040510 / Meta-Data Imported from CA Erwin v. 4 x / Meta-Data Imported from CA Erwin v. 4 x

(Attribute) Ship From Duns Number

Worthington / Legacy / ASN / 20040510 / Meta-Data Imported from CA Erwin v. 4 x / Meta-Data Imported from CA Erwin v. 4 x

(Attribute) Ship To Duns Number

Worthington / Legacy / ASN / 20040510 / Meta-Data Imported from Informatica / Meta-Data Imported from Informatica
Show standard vendor with SOX certification added
Mapping to other systems (impacted systems)
Legacy to ERP ETL Example

ERwin → ERwin XML → Meta-Data Repository → ERwin XML → ERwin

Informatica Source Analyzer

Informatica Target Analyzer

Informatica Source Analyzer

Informatica Target Analyzer

PowerCenter

Copyright © Meta Integration® Technologies, Inc.
Legacy to ERP ETL Example

- ERwin
- Meta-Data Repository
- Oracle ERP
- ASN
- OBDC
- ERwin XML
- ERwin XML
- Informatica Source Analyzer
- Informatica Target Analyzer
- ASN
- Oracle ERP

Copyright © Meta Integration® Technologies, Inc.

25 May 2005
Import From Informatica
Export into ERwin

Repository Manager

Export

Save
Manipulate Model in ERwin
Export back to Informatica as ETL Source
Model in MDR

---

[Diagram of a database model in a software interface, showing various attributes and their values, such as `Element Type: Attribute`, `Name: DUNS_NUMBER`, `ID: 125376`, `Physical ID: (405F911E-436)`, `Physical Name: DUNS_NUMBER`, `Status: Developmental`, etc.]

---

Meta Integration Works 4.1.0 Beta 2004/03/12
Subset ERP in ERwin
Oracle ERP Subset Model as Target of Informatica ETL
Standardization Support

- Standards Information Editing in ERwin
- ERwin XML
- Data Standards Packages in Meta-Data Repository
- ERwin XML
- Schema Forward Engineering Using ERwin
- OBDC
- XML/XSD

- Informatica
- Cognos

- Legacy Systems
- Oracle ERP
- Warehouse

25 May 2005

Copyright © Meta Integration® Technologies, Inc.
Standardization Example: Vendor Meta-Data Package

Meta Integration Works administrator (group Administrator) connected to localhost

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Name</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Physical ID</td>
<td>(DE7D36EE-C7B3-40D...</td>
</tr>
<tr>
<td></td>
<td>Physical Name</td>
<td>duns_nbr</td>
</tr>
<tr>
<td></td>
<td>Status</td>
<td>Developmental</td>
</tr>
<tr>
<td></td>
<td>Comment</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Design Level</td>
<td>Logical Physical</td>
</tr>
<tr>
<td></td>
<td>Stereotype</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Position</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Initial Value</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Optional</td>
<td>false</td>
</tr>
<tr>
<td></td>
<td>Cpp Scope</td>
<td>Public</td>
</tr>
<tr>
<td></td>
<td>Cpp Containment</td>
<td>Unspecified</td>
</tr>
<tr>
<td></td>
<td>Cpp Static</td>
<td>false</td>
</tr>
<tr>
<td></td>
<td>Cpp Derived</td>
<td>false</td>
</tr>
<tr>
<td></td>
<td>Constraint</td>
<td></td>
</tr>
</tbody>
</table>

Repository Manager: DBSvmst

Log Manager:
1. Loading the Directory from the database ... Please wait.
2. Downloading the Directory ... Please wait.
3. Retrieving the model from the server.
4. Loading the Model from the database ... Please wait.
5. Downloading the DBSvmst Model ... Please wait.
6. Ready...

1. Status

25 May 2005
Edit in ERwin
### Migrate Standard Reference Mappings

**Migration Wizard**

- **Migrate:** `\Worthington\Legacy\ASN\Vendor Standards Mapping\20040511 - SOX Update`
- **From:** `\Worthington\Legacy\ASN\Vendor Standards Mapping\20040510`

<table>
<thead>
<tr>
<th>Model Format</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>\Worthington\Legacy\ASN\20040510</code></td>
<td><code>\Worthington\Legacy\ASN\20040510</code></td>
</tr>
<tr>
<td><code>\Worthington\Standards\Vendor\SOX Version</code></td>
<td><code>\Worthington\Standards\Vendor\20040510</code></td>
</tr>
</tbody>
</table>

**Log Manager**

- Downloading the DBSvirt Model ... Please wait.
- Loading the Model from the database ... Please wait.
- Ready...

Copyright © Meta Integration Technologies, Inc.
Export to Other Meta-Data Environments

ERwin

Cognos

Informatica

XML, Rose, Java, Legacy, ETC.....
Conclusion