

Common Warehouse Metamodel (CWM), UML and XML

Dr. Daniel T. Chang

IBM Database Technology Institute
Chair, OMG CWM Working Group
(dtchang@us.ibm.com)

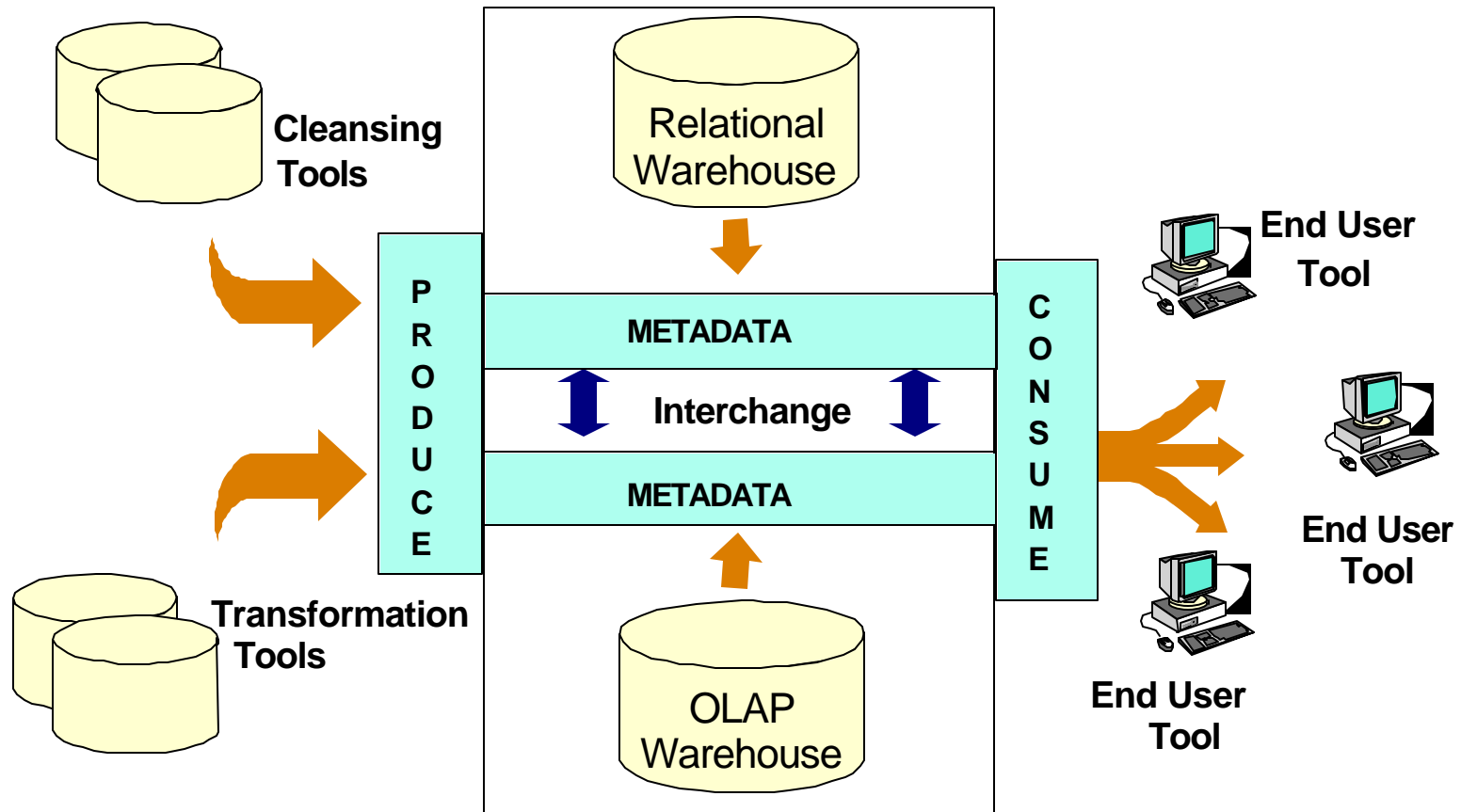
Meta Data Conference, March 19-23, 2000

Topics

- Why CWM?
- What is CWM?
- CWM and UML
- The CWM Metamodel
- CWM and XML
- CWM Extensions
- Conclusion

Why CWM?

The Problem Domain: Data Warehousing/Business Intelligence



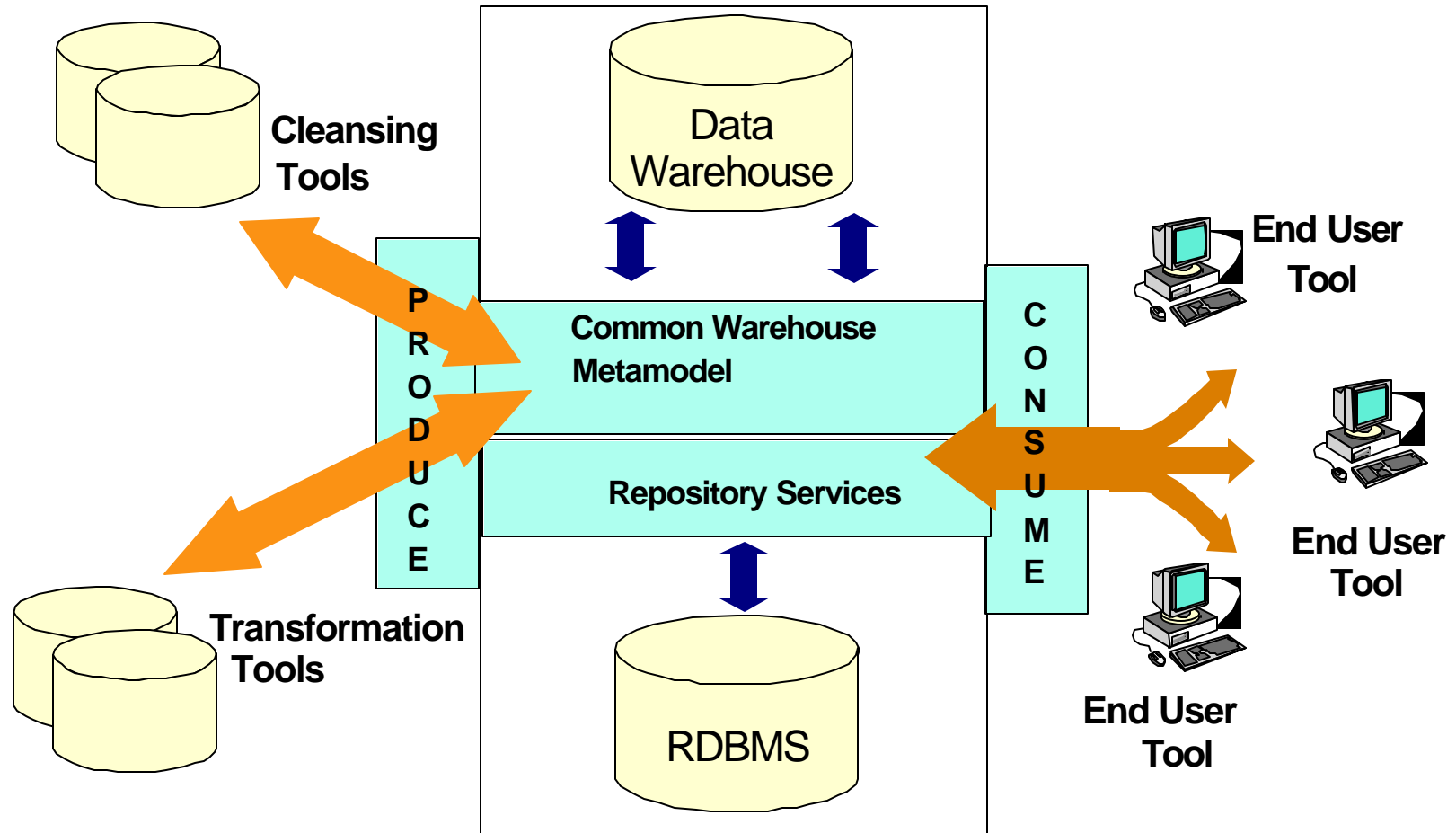
The Problem

- I/T Perspective
 - Product integration challenges are immense
 - End user requirements are continually changing
- User Perspective
 - We can't find the information that we need
 - The interpretation of information is a challenge
- Vendor Perspective
 - No industry standard way of sharing information
 - Metadata integration costs are significant

Problem Statement

- *Metadata management and integration* is the number one integration problem in data warehousing and business intelligence.
 - Data warehousing and business intelligence often involve the use of a variety of *tools and products*, each with its *own definition and format for metadata*.
 - *Creating, sharing and managing the metadata* for these tools and products is time consuming and error prone.

The Solution: CWM



Solution Framework

- A successful framework for solving the metadata management and integration problem must provide:
 - A **standard language** for defining the structure and semantics of metadata in a formal way
 - A **standard interchange mechanism** for sharing metadata defined in the standard language
 - A **common specification** that defines, in the standard language, the structure and semantics of shared metadata in data warehousing and business intelligence

Solution Requirements: CWMI RFP

- Interchange of all warehouse metadata including both **technical metadata** and **business metadata**
- Interchange of metadata that describes all **warehouse data elements** including data sources, transformations and data targets
- Interchange of metadata that describes all **warehouse processing elements** including scheduling, status reporting and history recording

Solution Requirements: CWMI RFP

- Interchange of metadata that describes **informational data** and the use of major types of **informational data models** (such as multidimensional and hierarchical classification) for representing informational data
- Interchange of metadata that describes **operational data** and the use of major types of **operational data models** (such as relational, object-oriented and hierarchical) for representing operational data

What is CWM?

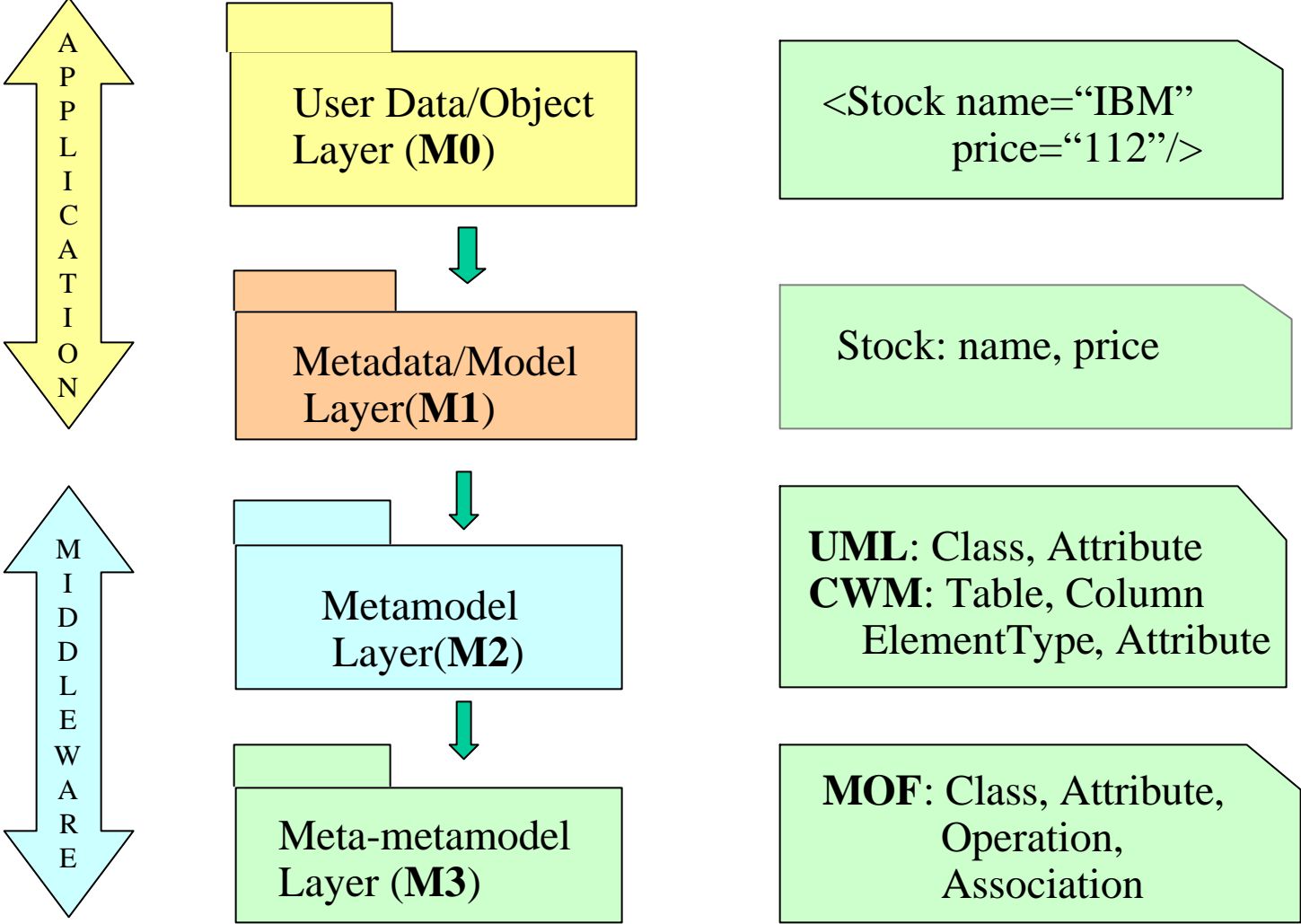
CWM

- A complete specification of the **syntax and semantics** needed to **export/import** shared warehouse metadata and the common warehouse metamodel, including:
 - **The CWM Metamodel** (Volume 1)
 - Interchange format for shared warehouse metadata (**CWM DTD**, Volume 2)
 - Interchange format for the CWM Metamodel (**CWM XML**, Volume 2)
 - Access API for shared warehouse metadata (**CWM IDL**, Volume 2)

CWM Design Basis (I)

- **OMG Metamodeling Architecture**
The best starting point for developing a solution framework
 - Metamodeling language (**M3**)
 - Metamodels (**M2**)
 - Metadata or models (**M1**)
 - Data or objects (**M0**)

OMG Metamodeling Architecture



OMG Modeling Architecture

Applications, Tools, Repositories

Metamodels (UML, CWM, ...)

Meta Object Facility (MOF Model, MOF-IDL)
XML Metadata Interchange (XMI)

CWM Co-submitting Companies

- **IBM** (*Dan Chang, J. J. Daudenarde, Debra LaVergne, Christoph Lingenfelder*)
- **Unisys** (*Sridhar Iyengar, Don Baisley, Doug Tolbert*)
- **NCR** (*Vihelm Rosenqvist, Bruce McLean*)
- **Hyperion** (*John Poole, David Zhang*)
- **Oracle** (*David Last, David Mellor, Mark Hornick*)
- **UBS** (*Hans-Peter Hoidn, Jeffrey Peckham*)
- **Genesis** (*David Frankel, Phil Longden*)
- **Dimension EDI** (*Chris Nelson, Anders Tornqvist*)

- Expertise in UML, XML, metadata repository, databases, data warehousing, and business intelligence (OLAP, data mining)

CWM Supporting Companies

- **Deere** (*Dave Smith*)
- **Sun** (*Chuck Mosher, Karsten Riemer, Nidhi Rao*)
- **HP** (*Jishnu Mukerji*)
- **Data Access** (*Cory Casanave*)
- **InLine Software** (*Jack Greenfield*)
- **Aonix** (*Charles Simon*)
- **Hitachi** (*Yuichi Yugawa*)

- Expertise in using databases, data warehouses, and business intelligence tools

CWM and UML

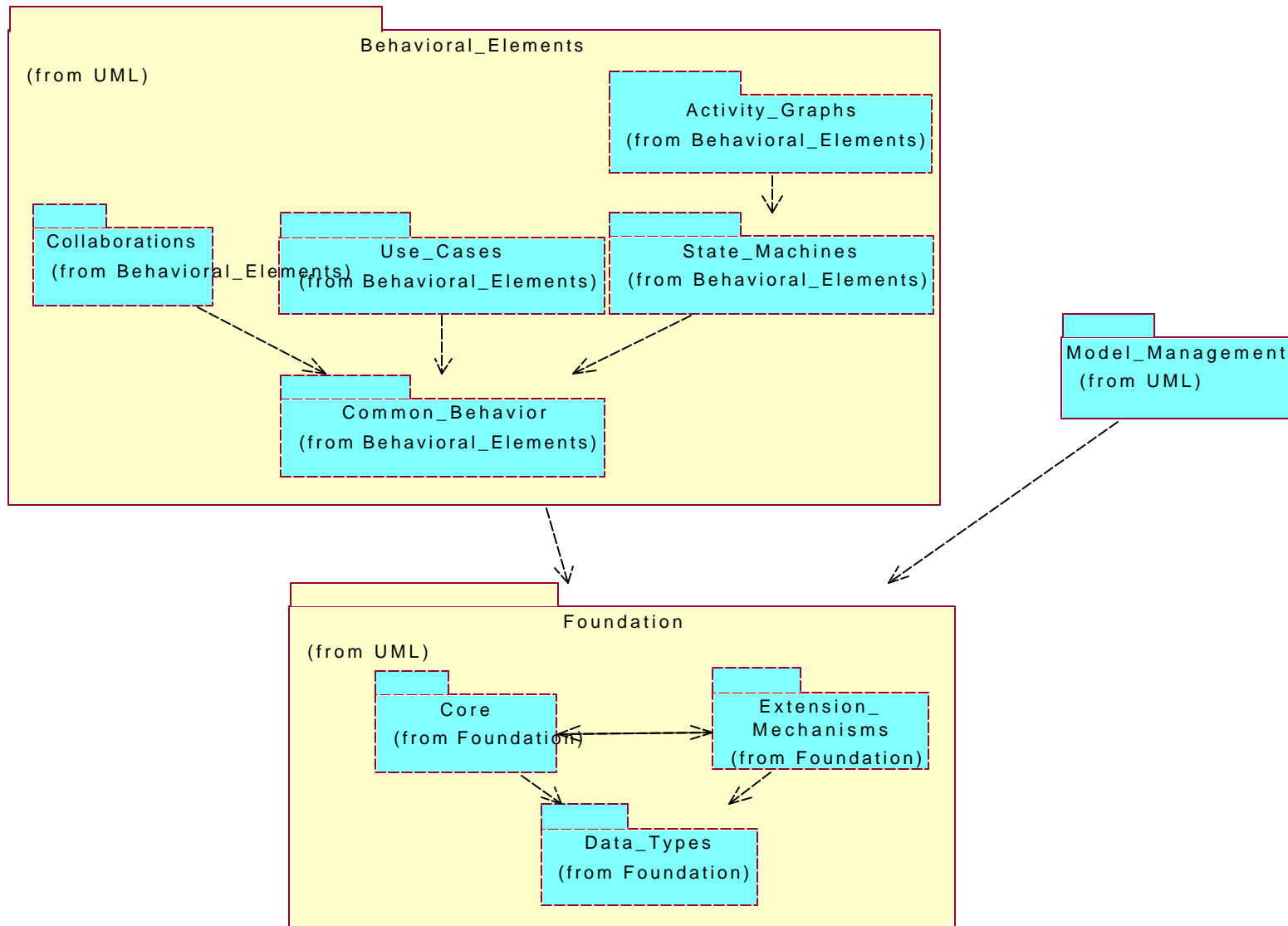
Why UML - Interchange?

- Why not XML DTD or XML Schema?
 - XML DTD
 - Primitive data model: no inheritance, no operations, no associations, no constraints
 - Only string data types
 - XML Schema
 - Same primitive data model as above
 - Richer data types
- **UML**
 - Rich object-oriented model with associations/constraints
 - Extensible data types: CORBA (MOF), etc.
 - **UML => XML DTD (per XMI) and/or XML Schema (per XMI, coming soon)**

Why UML- Access?

- Why not CORBA IDL or Java?
 - CORBA IDL
 - Interface only, little structure or semantics
 - CORBA data types
 - Java
 - Java object model
 - Java data types
- **UML**
 - Rich object-oriented model with associations/constraints
 - Extensible data types: CORBA (MOF), etc.
 - **UML => CORBA IDL (per MOF) and/or Java (per JMI, coming soon)**

UML 1.3



CWM Design Basis (II)

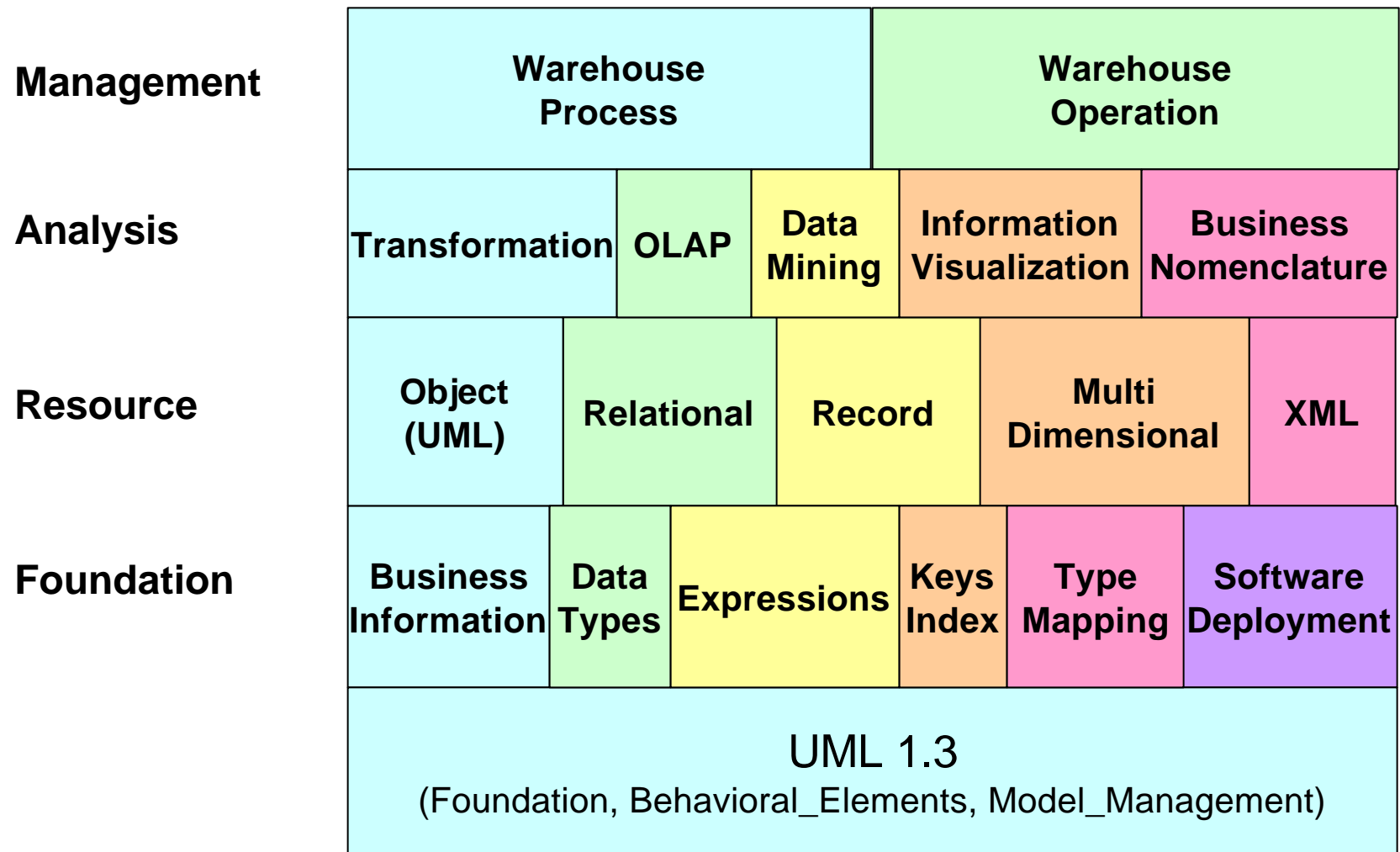
- **OMG Metamodeling Architecture**
 - **UML** as the standard language for defining models of metadata
 - **UML semantics** (Class, Attribute, Operation; Association, Role/AssociationEnd; Constraint)
 - **UML notation** (class diagram, object diagram, collaboration diagram)
 - **OCL** (Object Constraint Language)

Roles of UML in CWM

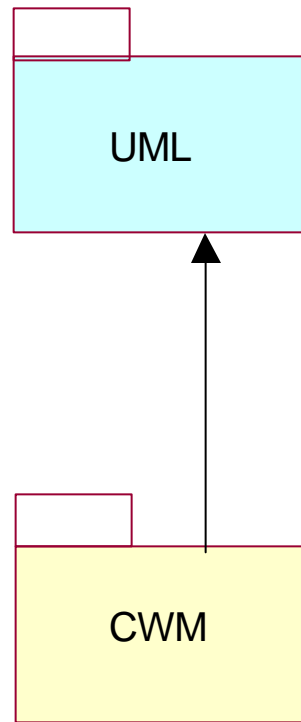
- The MOF-equivalent meta-metamodel
 - UML Semantics, UML Notation, OCL
- The foundation metamodel
 - UML Foundation, Common_Behavior, and Model_Management packages
- The object (resource) metamodel
 - Same as above

The CWM Metamodel

The CWM Metamodel

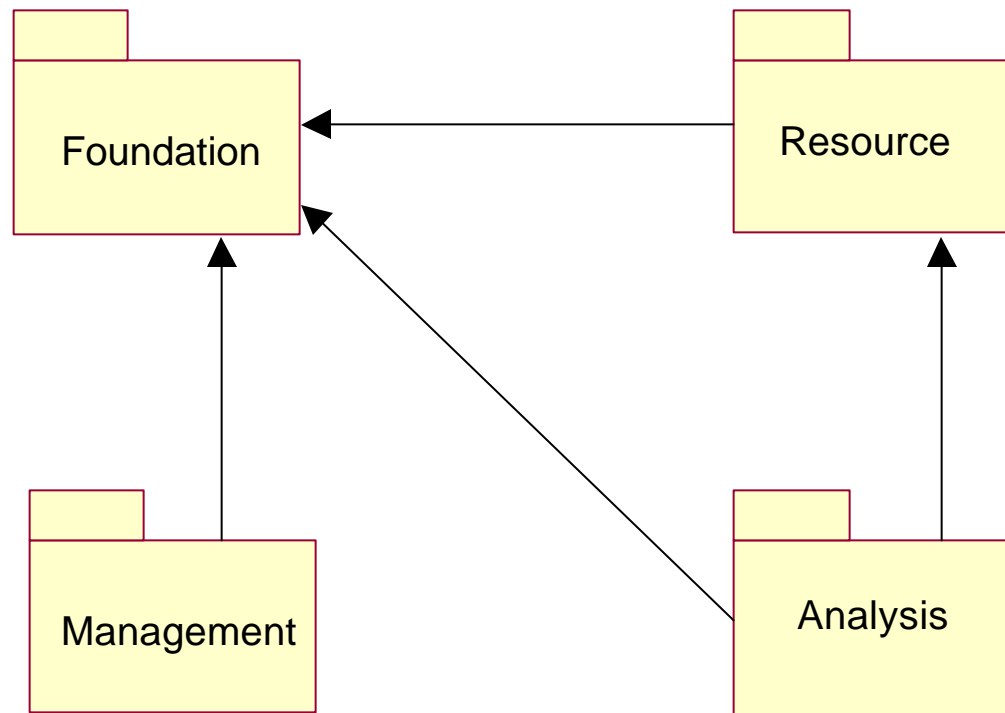


CWM - Top Level

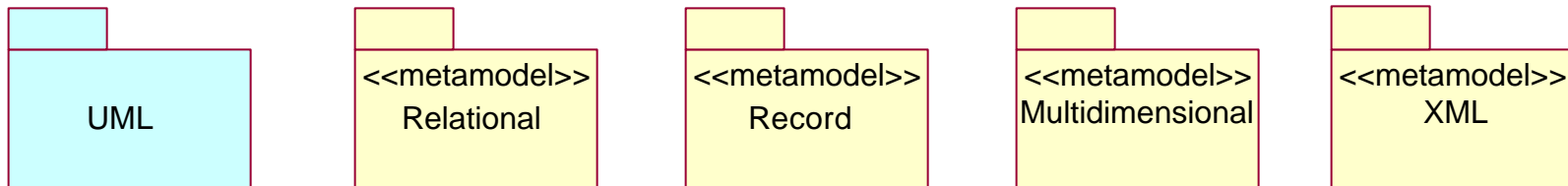


The top-level packages in CWM:
org.omg.uml { UML 1.3 }
org.omg.cwm

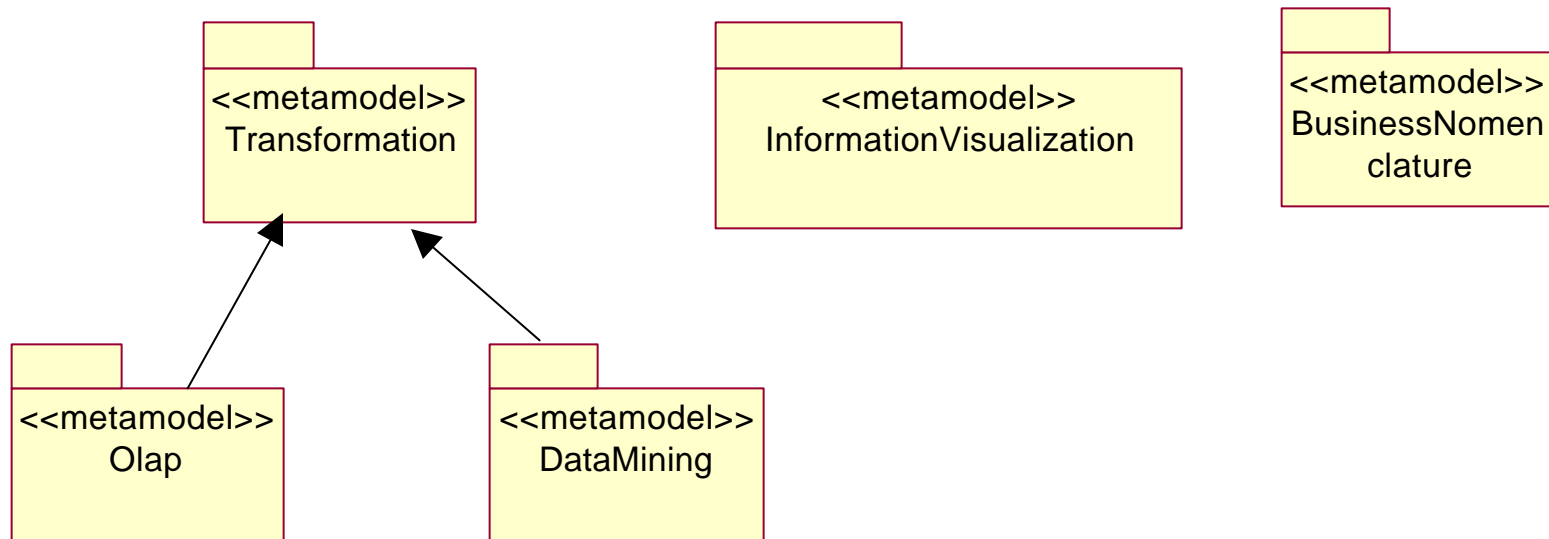
CWM - Overview



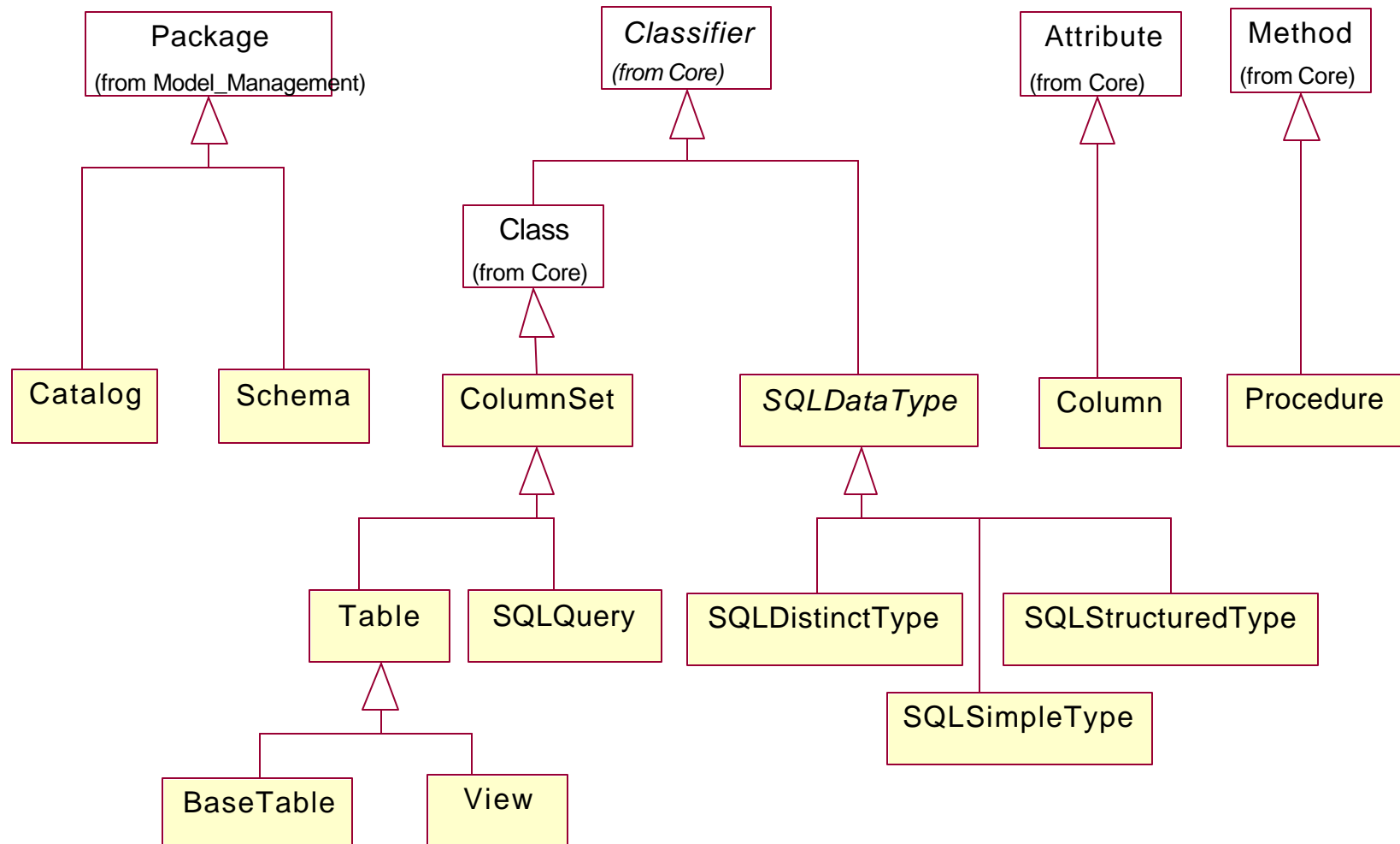
CWM - Resource



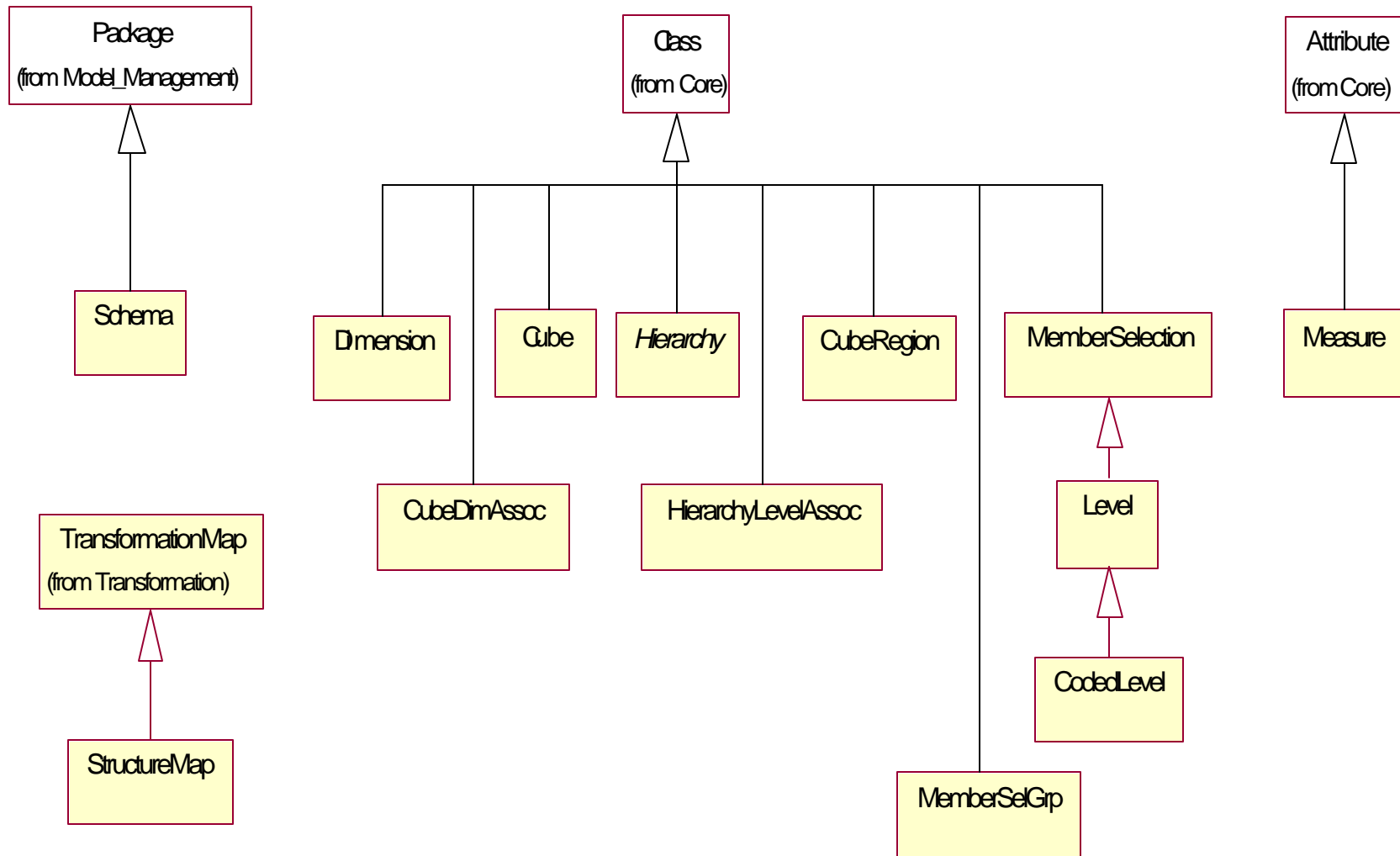
CWM - Analysis



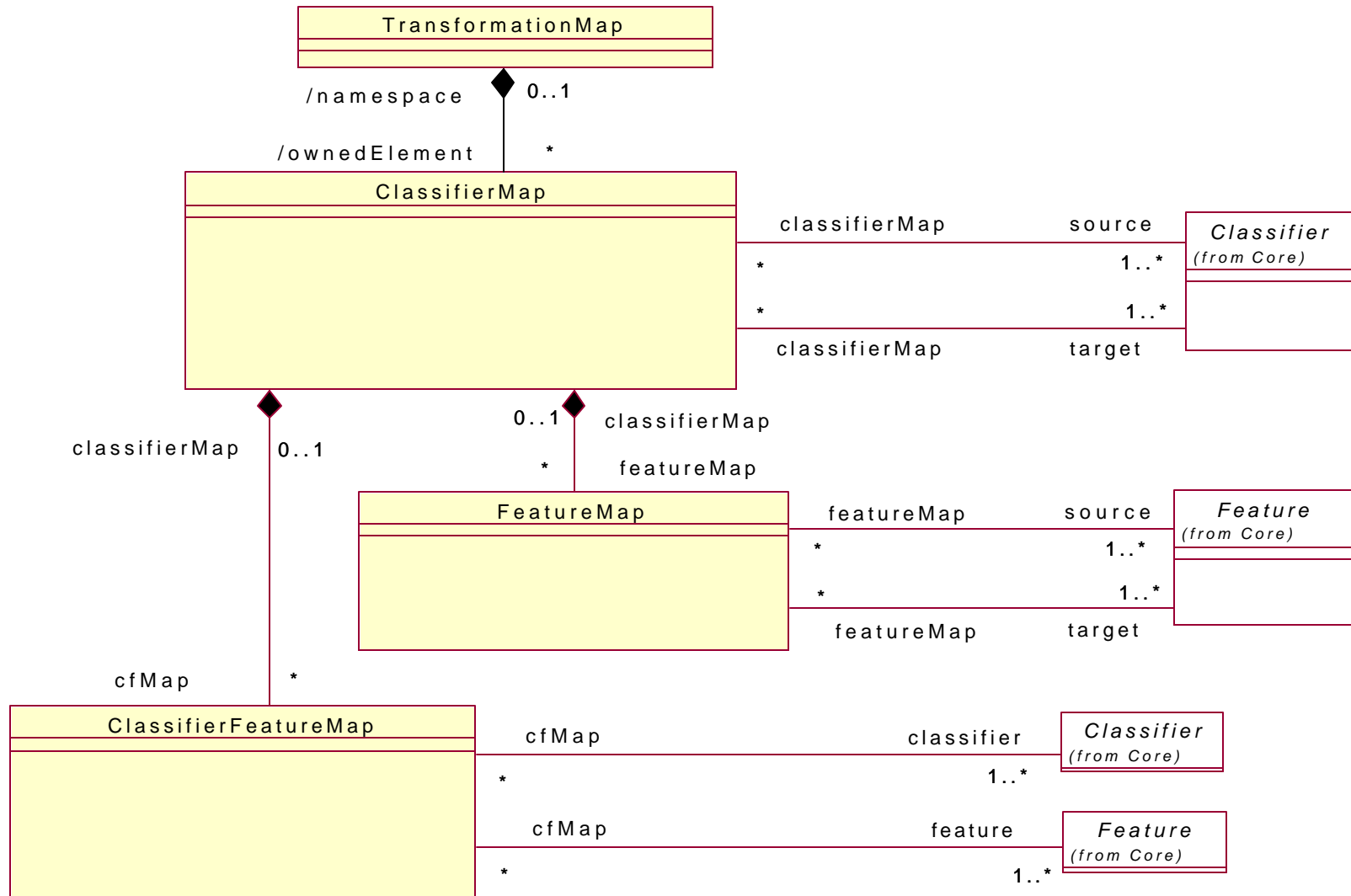
Relational Metamodel



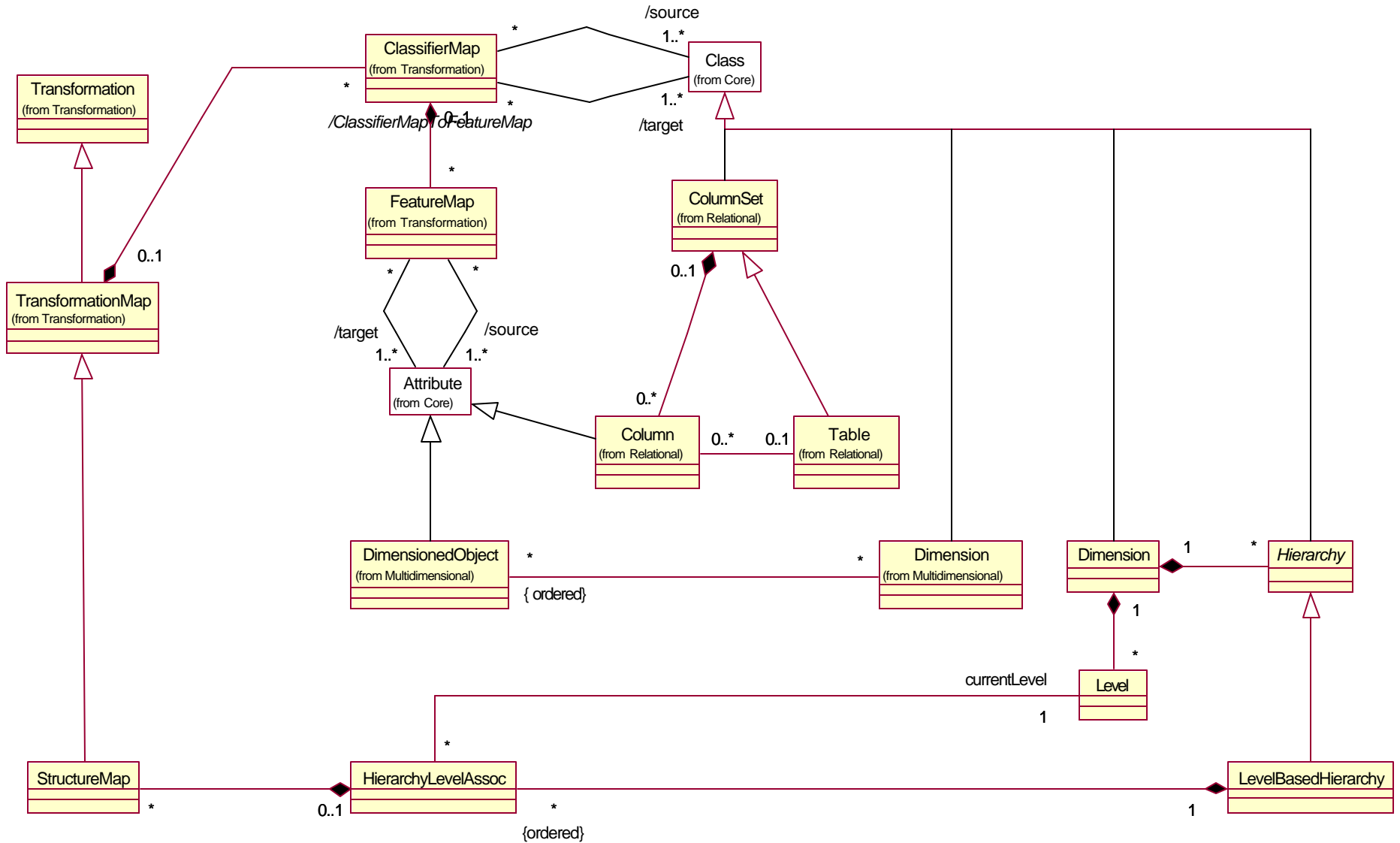
OLAP Metamodel



Transformation Metamodel



OLAP Deployment to Relational

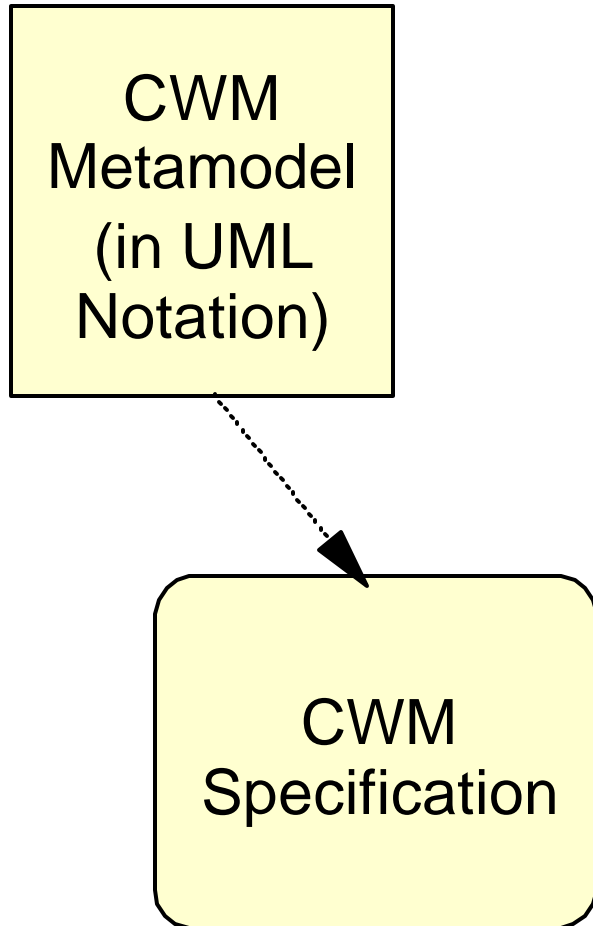


Dan Chang

CWM Resource/Analysis Matrix

UML	Package	Class	Attribute		
Object	Package	Class	Attribute		
Relational	Catalog	Schema	Table	Column	
Record	RecordFile	RecordDef	Field		
Multidimensional	Schema	Dimension	DimensionedObject		
XML	Schema	ElementType	Attribute		
OLAP	Schema	Dimension	Cube	Attribute	Measure

CWM Specification

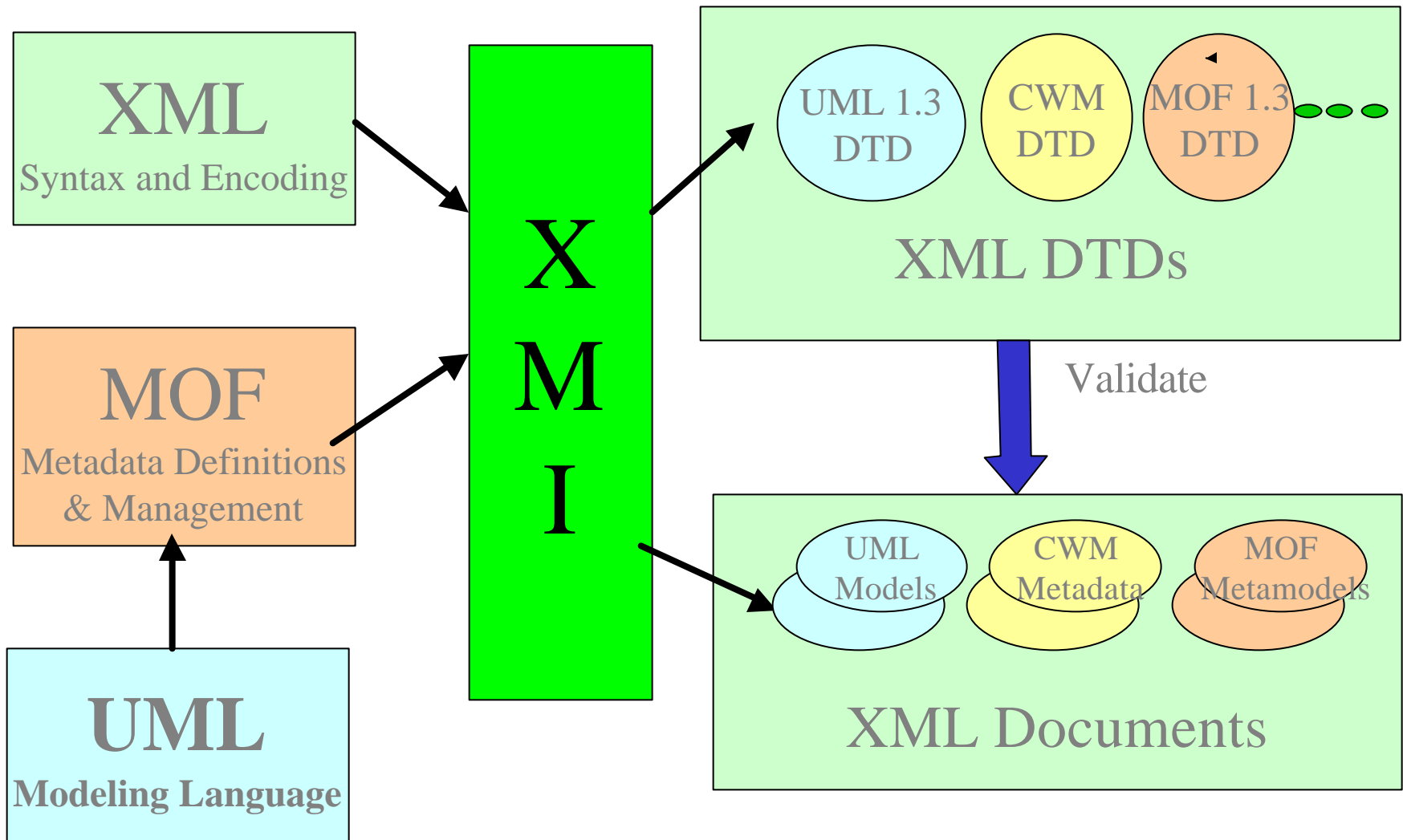


CWM and XML

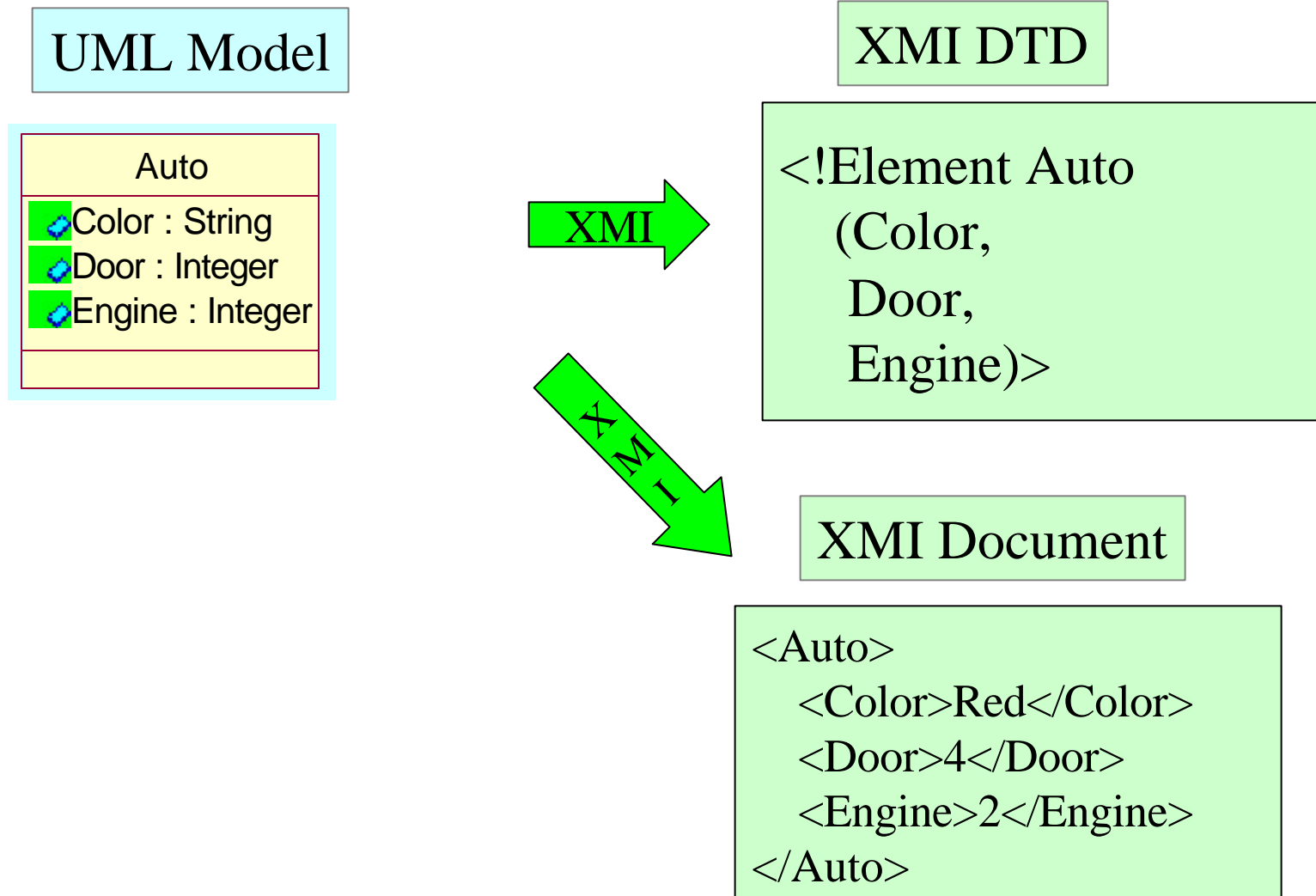
XML Metadata Interchange (XMI)

- Use **W3C XML** for the transfer syntax and interchange format
 - Specify **XML DTD** to enable transfer and verification of
 - **MOF-based metamodels (using MOF DTD)**
 - **UML-based models (using UML DTD)**
- Specify a precise **MOF to XML Mapping**
 - Enables automatic generation of **XML DTDs**
 - Enables automatic generation of **XML documents**

XMI Simplified



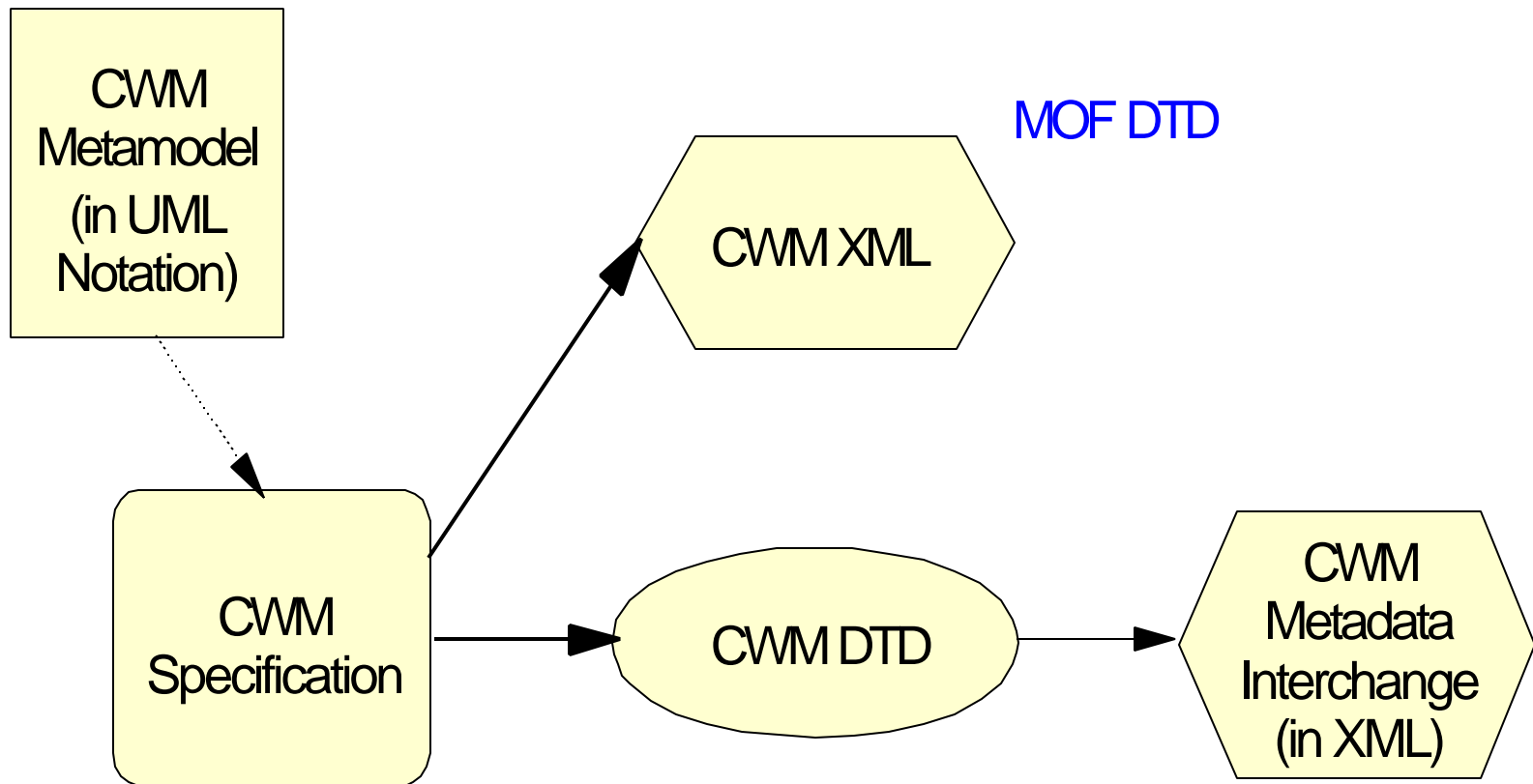
XMI - Automobile Example



CWM Design Basis (III)

- **OMG Metamodeling Architecture**
 - **UML** as the standard language for defining models of metadata
 - **XMI** as the standard mechanism for interchanging metadata and metamodels in **XML**
 - **XML DTD** Production Rules
 - **XML Document** Production Rules

CWM Specification: CWM XML, CWM DTD

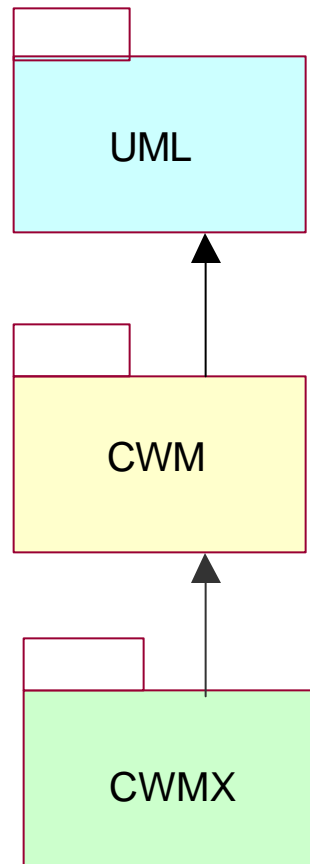


CWM Extensions

CWM Extensions (CWMX)

- Published vendor specific metamodel for the purpose of metadata interchange (Volume 3)
- Common ancestry in the CWM metamodel
- Demonstrates the validity of the CWM metamodel
- Demonstrates the extensibility of the CWM metamodel

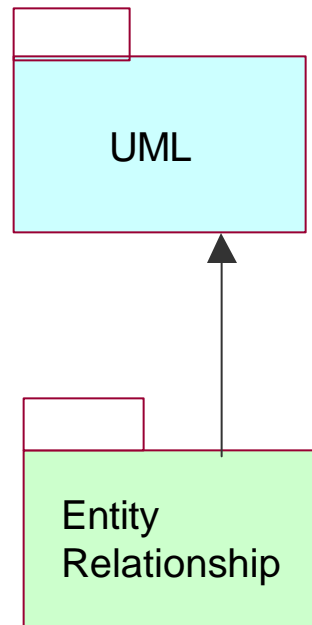
CWMX - Top Level



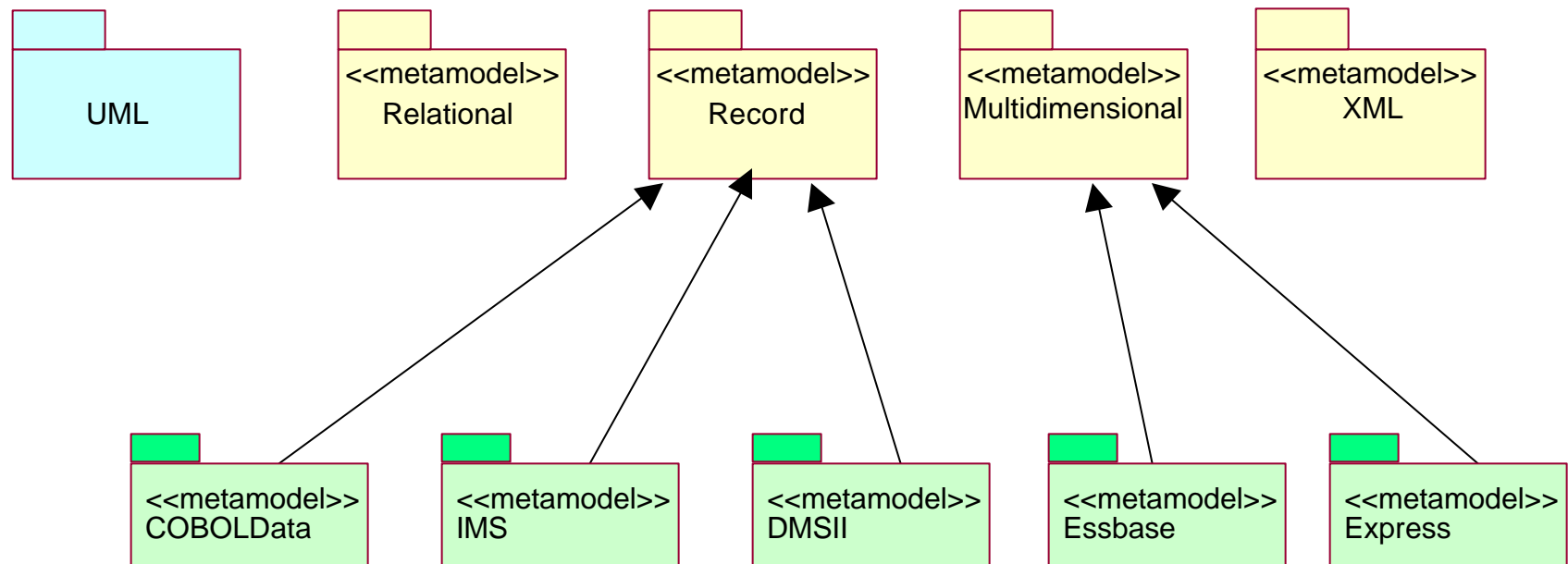
The top-level packages in CWMX:

org.omg.uml { UML 1.3 }
org.omg.cwm
org.omg.cwmx

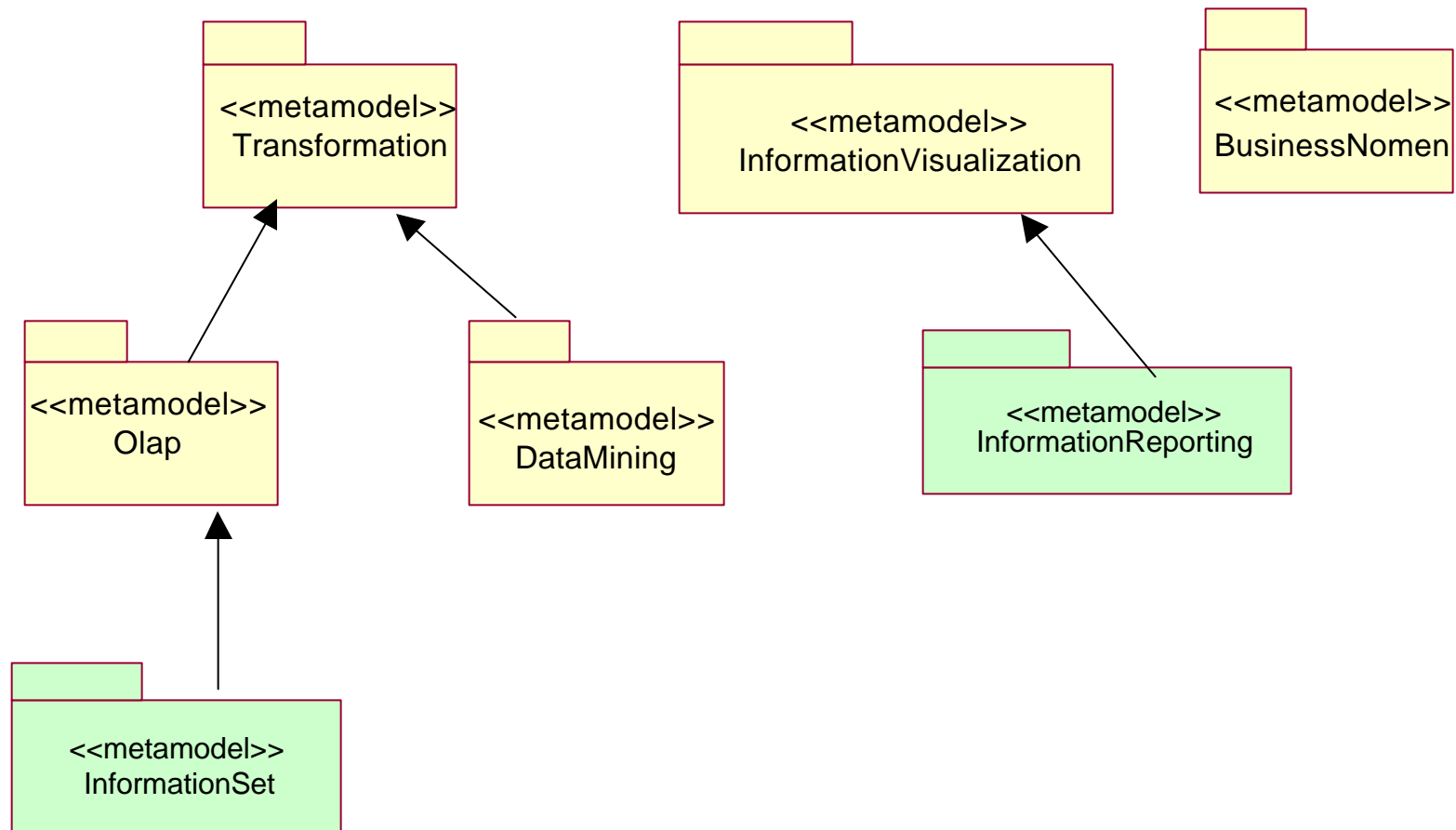
CWMX - Foundation



CWMX - Resource



CWMX - Analysis



Conclusion

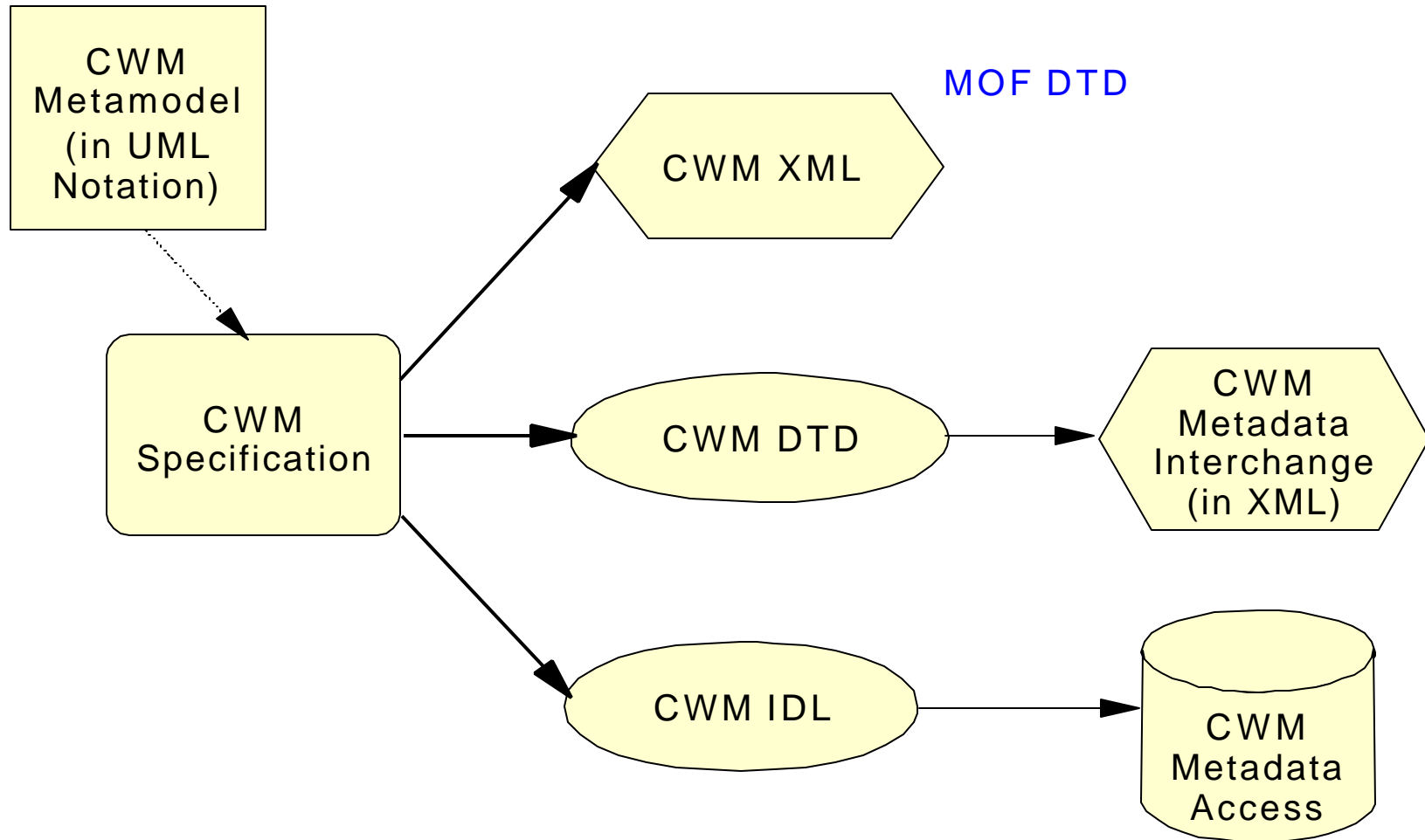
CWM Design Basis

- **OMG Metamodeling Architecture**
 - **UML** as the standard language for defining metamodels
 - **XMI** as the standard mechanism for interchanging metadata and metamodels in **XML**
 - **MOF to IDL Mapping** as the standard mechanism for accessing metadata through APIs (independent of programming languages and object models)
 - **MOF to Java Mapping** as the standard mechanism for accessing metadata in Java (coming soon)

CWM

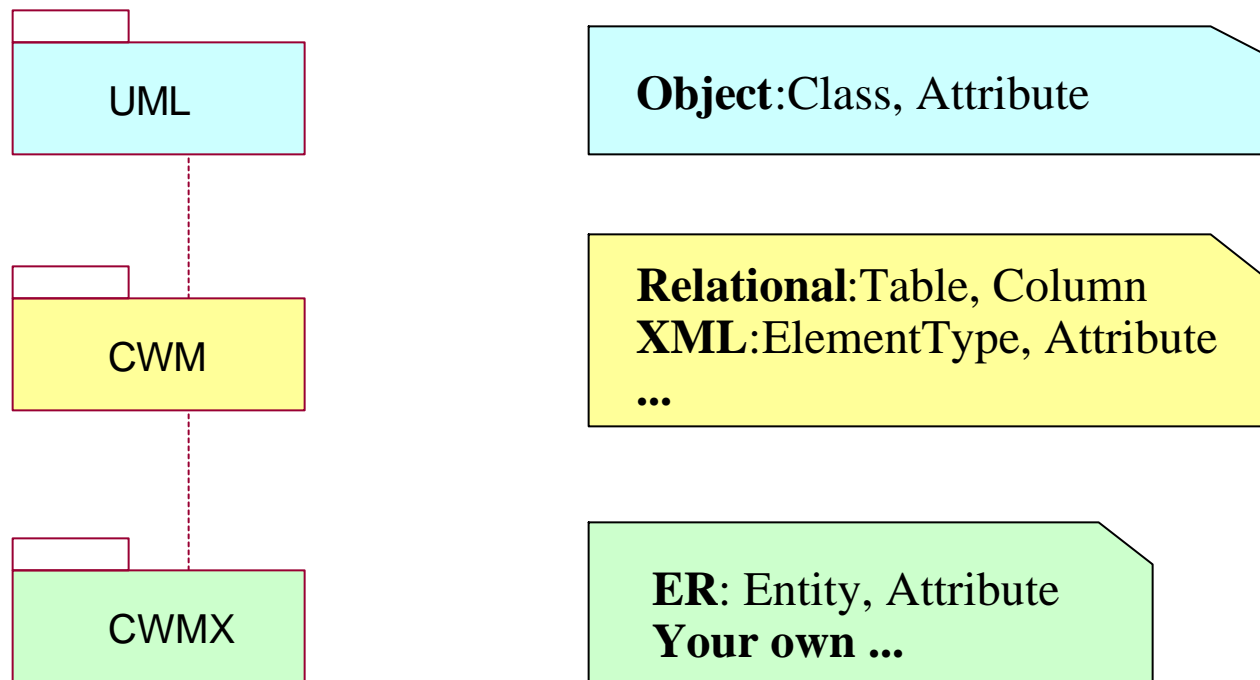
- A **common specification** that defines, in **UML**, the structure and semantics of shared metadata in **data warehousing and business intelligence**
 - **Resources**: Object, Relational, Record, Multidimensional, XML
 - **Analysis**: Transformation, OLAP, Data Mining, Information Visualization, Business Nomenclature
- A **common specification** that defines, in **XML**, the interchange format and, in **IDL**, the access API of such shared metadata

CWM Specification: CWM XML, CWM DTD, CWM IDL



CWM

- Enables interchange and access of shared metadata at *three abstraction levels*



CWM: Past, Present and Future

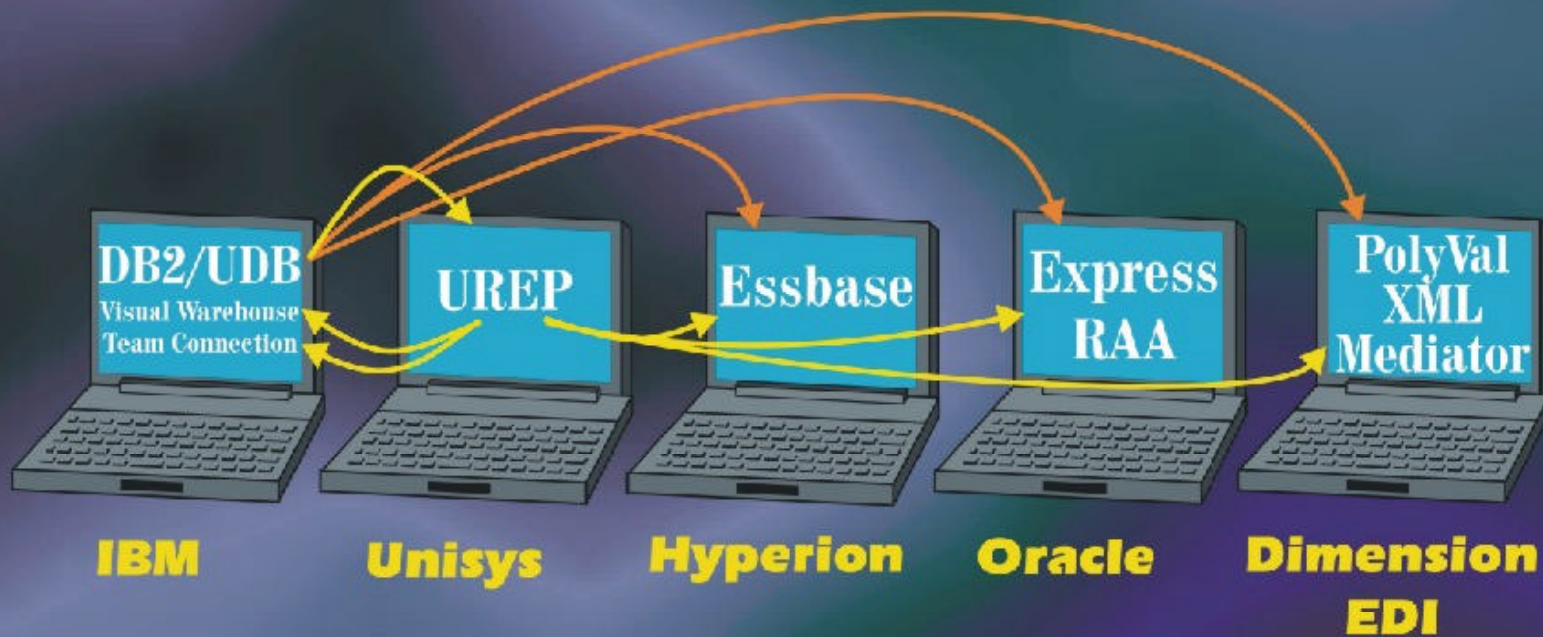
- Past
 - Initial submission: 9/17/99
 - OMG Demo: 11/99
 - Evaluation: 9/99 - 1/00
- Present
 - Final submission: 2/11/00
 - Evaluation: 2/00 - present
- Future
 - Adopted specification: 3/00 or 6/00
 - Available specification: 9/00 or 12/00
 - Product enablement and interoperability showcase

November 1999 OMG Demo

Common Warehouse Metamodel

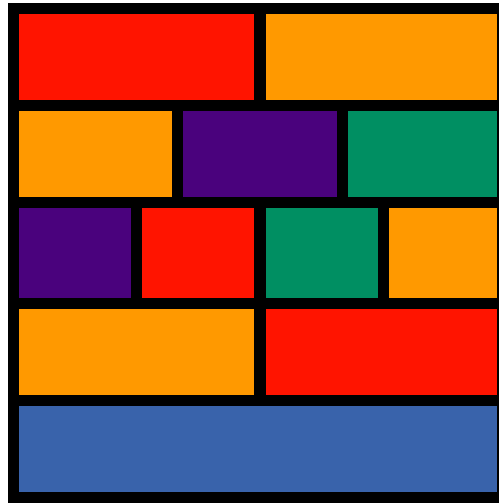
t e c h n o l o g y d e m o

Metadata Interchange Flow



CWM References

- **OMG CWMI RFP Web page**
http://www.omg.org/techprocess/meetings/schedule/CWMI_RFP.html
 - CWM Specification (ad/2000-01-01)
 - CWM Specification Volume 2. XML, IDL and DTD (ad/2000-01-02)
 - CWM Specification Volume 3. Extensions (CWMX) (ad/2000-01-03)
 - CWM Specification Volume 4. Extensions XML, IDL and DTD (ad/2000-01-11)
- **CWM Forum Web site**
<http://www.cwmforum.org/>



C W M

common warehouse metamodel

Dan Chang