



Jan. 15, 2008

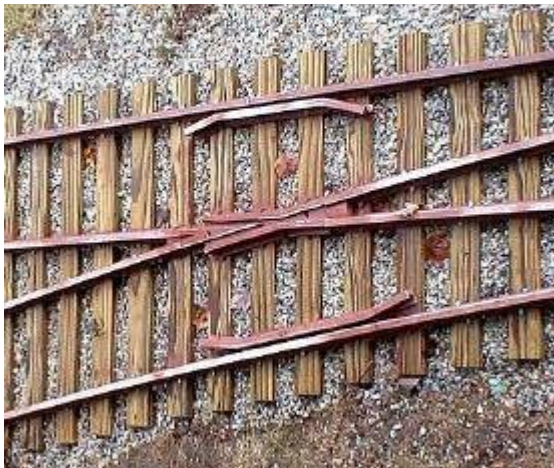
“Business Metadata”

In Search of Meaning in Our Data

Donald J. Soulsby
metaWright.com

METAWRIGHT

Communication



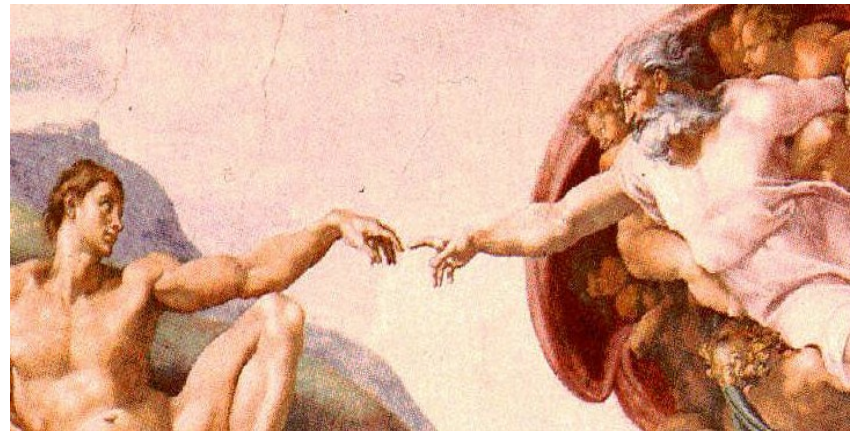
Agenda

- **Metadata Lifecycle**
- **Metadata Navigation**
- **DW 2.0 Business Metadata**
- **Data Rationalization**
- **Semantics, Ontologies & Taxonomies**

Electronic Data Processing Era

- **The “Filing Problem”**
 - Focus on process & delivery of data collection systems
 - Application Development tools (languages)

**Data
Creation**



METAWRIGHT

“You don’t have a *filing* problem,
You have a *finding* problem”

METAWRIGHT

Information Systems Era

- **The “Finding” problem**
 - Data Warehouse, Business Intelligence, Metadata Management
 - ETL, Reporting Tools, Data Modeling, Repository

**Data
Archaeology**



METAWRIGHT

Data Resource Management Standards

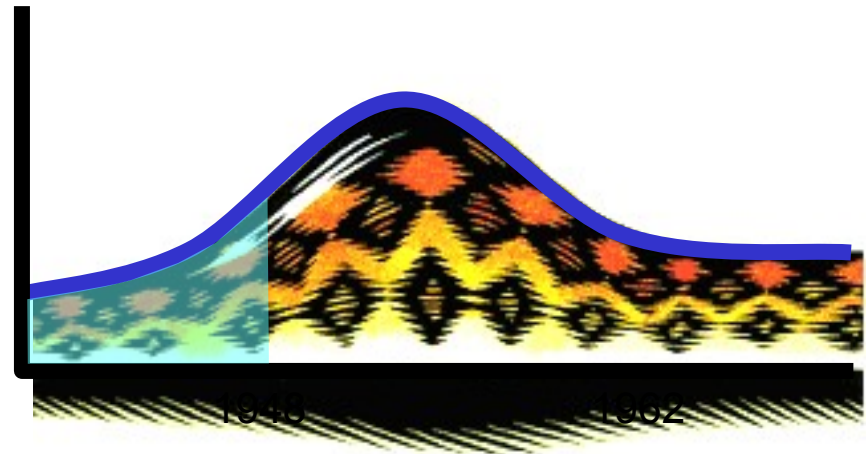
- **Data Models**
- **Data Stewards**
- **Naming standards**
- **Business Definitions**

Bye Bye Boomers

85

Age + Years of
Service

The Pig and the Python



“The Pig and the Python:
How to Prosper from the Aging Baby Boom”
by David Cork, Susan Lightstone

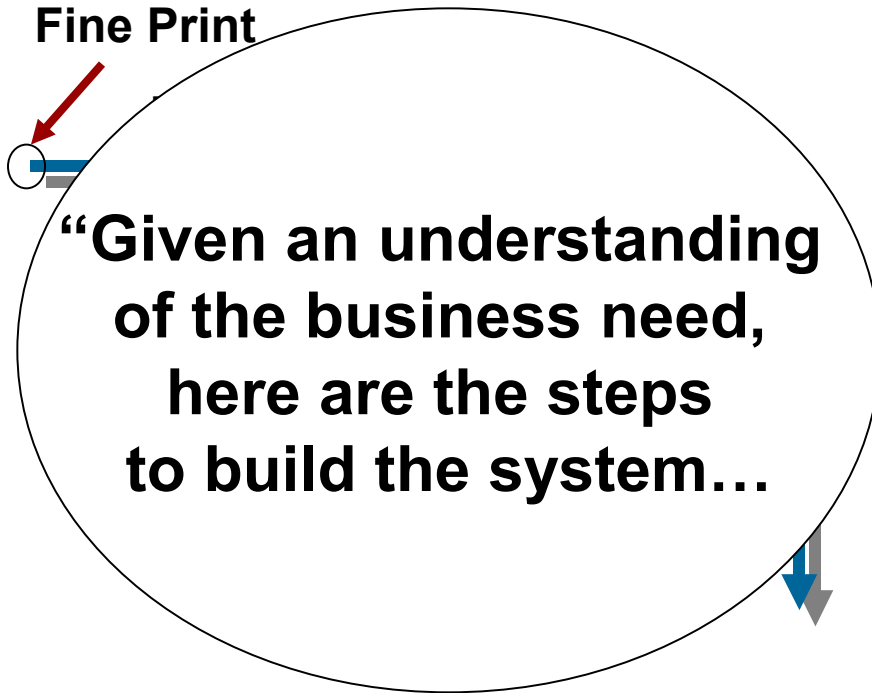
Pig in the Python – (1946-1964)

- **AARP study, more than 60 percent of U.S. companies are bringing back retirees as contractors or consultants.**

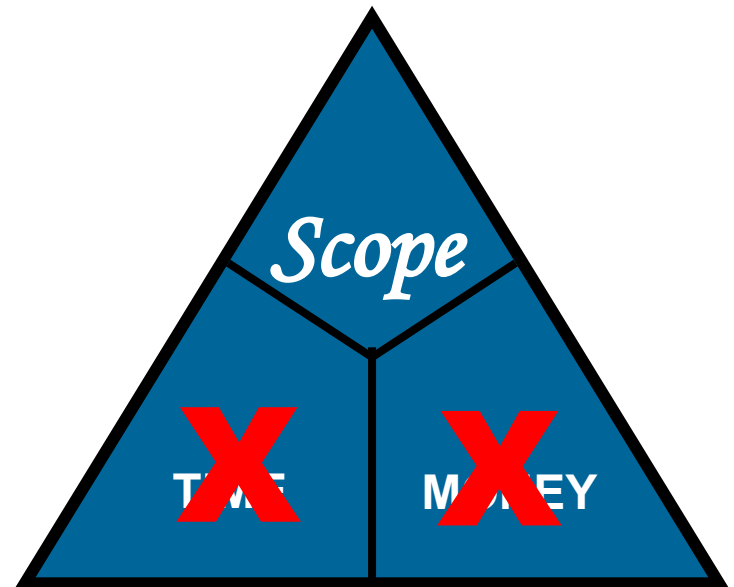
Montenegro, Xenia, Linda Fisher, and Shereen Remez (2002). “Staying Ahead of the Curve: The AARP Work and Career Study (A National Survey Conducted for AARP by RoperASW), Washington, D.C., September.

Lack of Integration - Interoperability

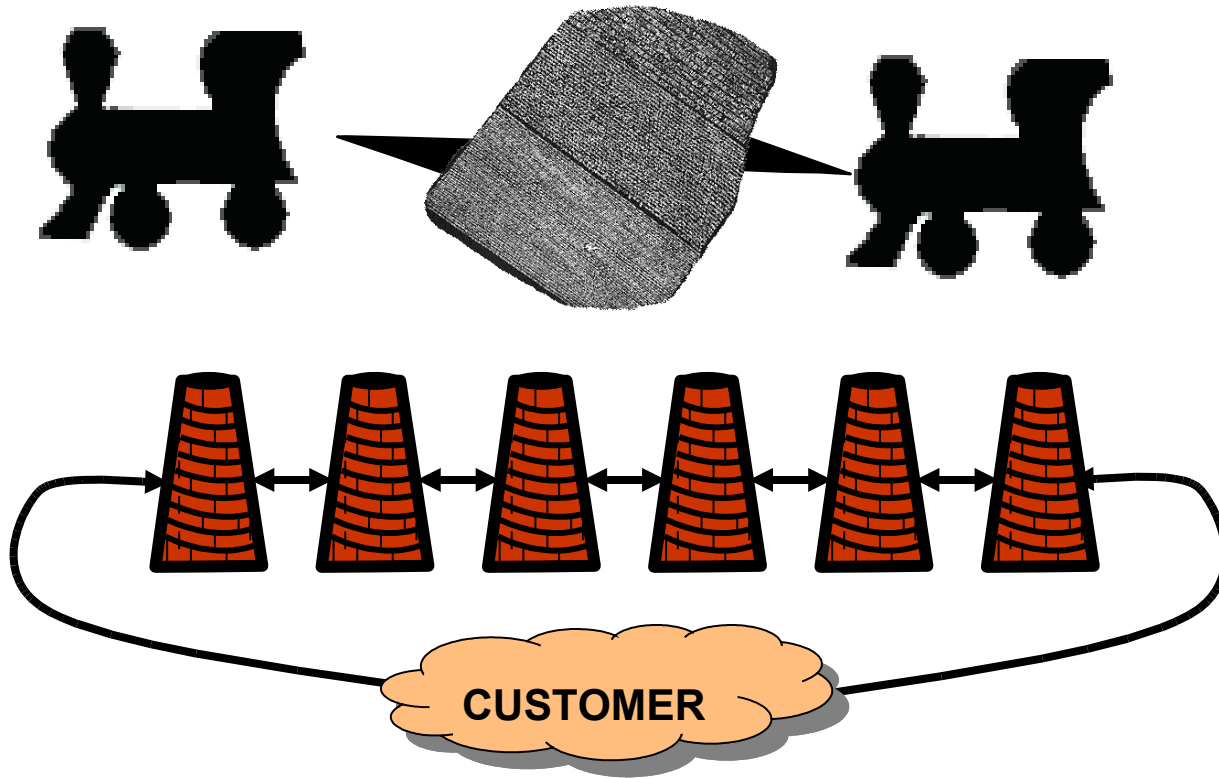
Fine Print



Waterfall Approach



Enterprise Data Model



META WRIGHT

Metadata Historical Perspective

- **Not Partnering with the business**
- **Lack of management understanding and expectations**
- **Lack of meaningful business case**
- **Taking short-sighted or technology based approaches**
- **Not useful, WORN**
- **Metadata as a stimulus to a Business Event will be number 11 on a list of 10**

The *Integrated* and *Interoperable* Age!

METAWRIGHT

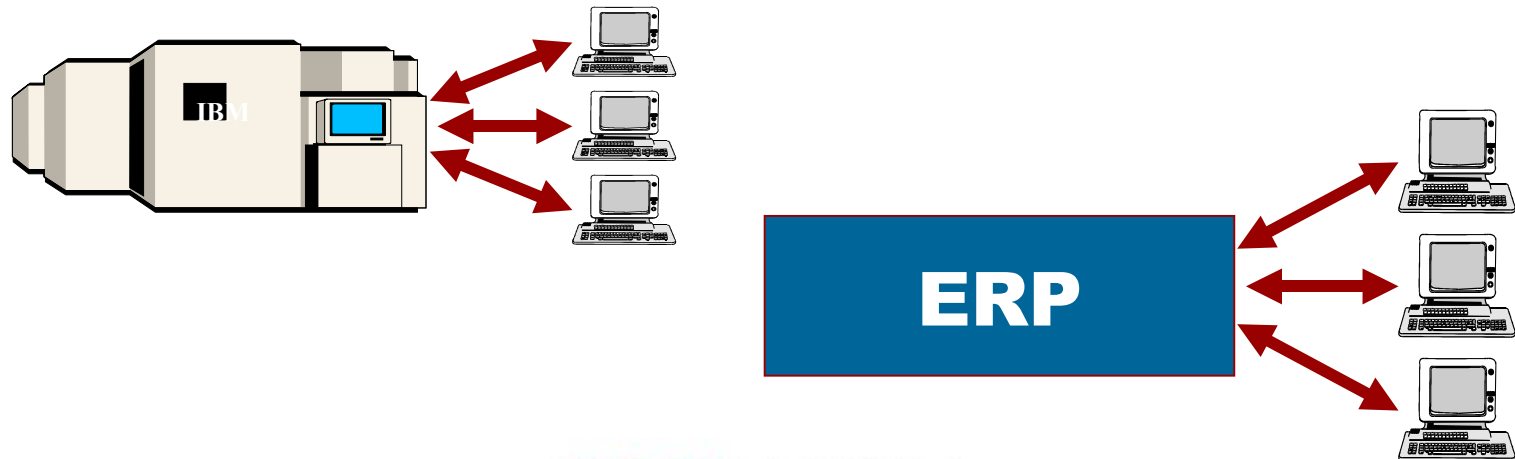
Business Service Management

- **An approach to the management of IT services that considers the business processes supported and the business value provided.**
- **A CIO Magazine – CA BSM Survey**
 - **IT Priorities (4-5 Critical Priority)**
 - Align IT with Business Priorities (87%)
 - Improve Service to End Users (75%)
 - Controlling IT Cost (75%)

http://ca.com/files/WhitePapers/ca_cio2cio_v2exec_summary.pdf

Customer Data Integration (CDI)

- **The technology, processes and services needed to create and maintain an accurate, timely and complete & comprehensive representation of a customer across multiple channels, business lines, and enterprises.**



METAWRIGHT

Metadata Lifecycle

*META*WRIGHT

Stewardship

- **Not a Single DATA Steward**

- Business
Technical
Operational
- Accountability
Authority
Responsibility

- **Form vs.
Content**

		Columns		
Rows		Cells		

Metadata Domains

Zachman Framework for Enterprise Architecture

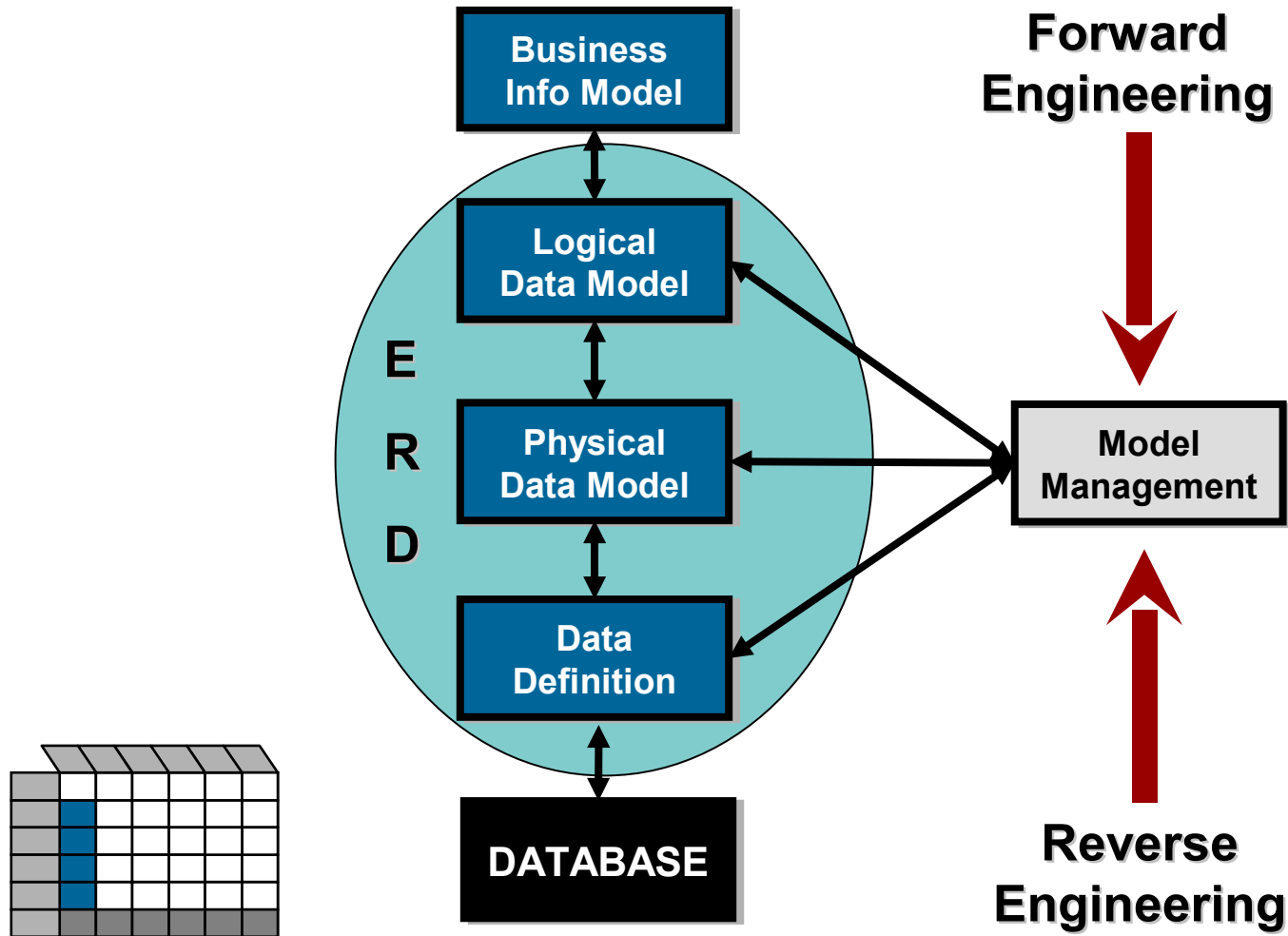
	WHAT	HOW	WHERE	WHO	WHEN	WHY	
CONTEXTUAL Scope	List of things	List of Processes	List of Locations	List of Organization Units	List of Events	List of Business Goals/Stat.	Business
CONCEPTUAL Business Model	Business Entity Model	Business Process Model	Logistics Network	Work Flow Model	Master Schedule	Business Plan	
LOGICAL System Model	Logical Data Model	Application Process Model	System Network Model	Human Interface Paradigm	Processing Structure	Business Rule Model	
PHYSICAL Technology Model	Physical Data Model	Application Structure Chart	Network Technology Model	Presentation Architecture	Control Structure	Rule Design	Technical
OUT-OF-CONTEXT Components	Data Definition	Program	Network Components	Interface Components	Timing Definition	Rule Specification	
PRODUCT Functioning System	DATABASE	APPLICATION	NETWORK	ORGANIZATION	SCHEDULE	STRATEGY	

Enterprise Metadata Integration

Zachman Framework for Enterprise Architecture

	WHAT	HOW	WHERE	WHO	WHEN	WHY
CONTEXTUAL Scope	List of things	List of Processes	List of Locations	List of Organization Units	List of Events	List of Business Goals/Stat.
CONCEPTUAL Business Model	Business Entity Model	Business Process Model	Logistics Network	Work Flow Model	Master Schedule	Business Plan
LOGICAL System Model	Logical Data Model	Business Process Model	System Network Model	Human Interface Paradigm	Processing Structure	Business Rule Model
PHYSICAL Technology Model	Physical Data Model	Business Process Model	Network Technology Model	Presentation Architecture	Control Structure	Rule Design
OUT-OF-CONTEXT Components	Data Definition	Business Process Model	Network Components	Interface Components	Timing Definition	Rule Specification
PRODUCT Functioning System		OPERATION	NETWORK	ORGANIZATION	SCHEDULE	STRATEGY

Data Management Process



William H. Inmon – Technical Metadata

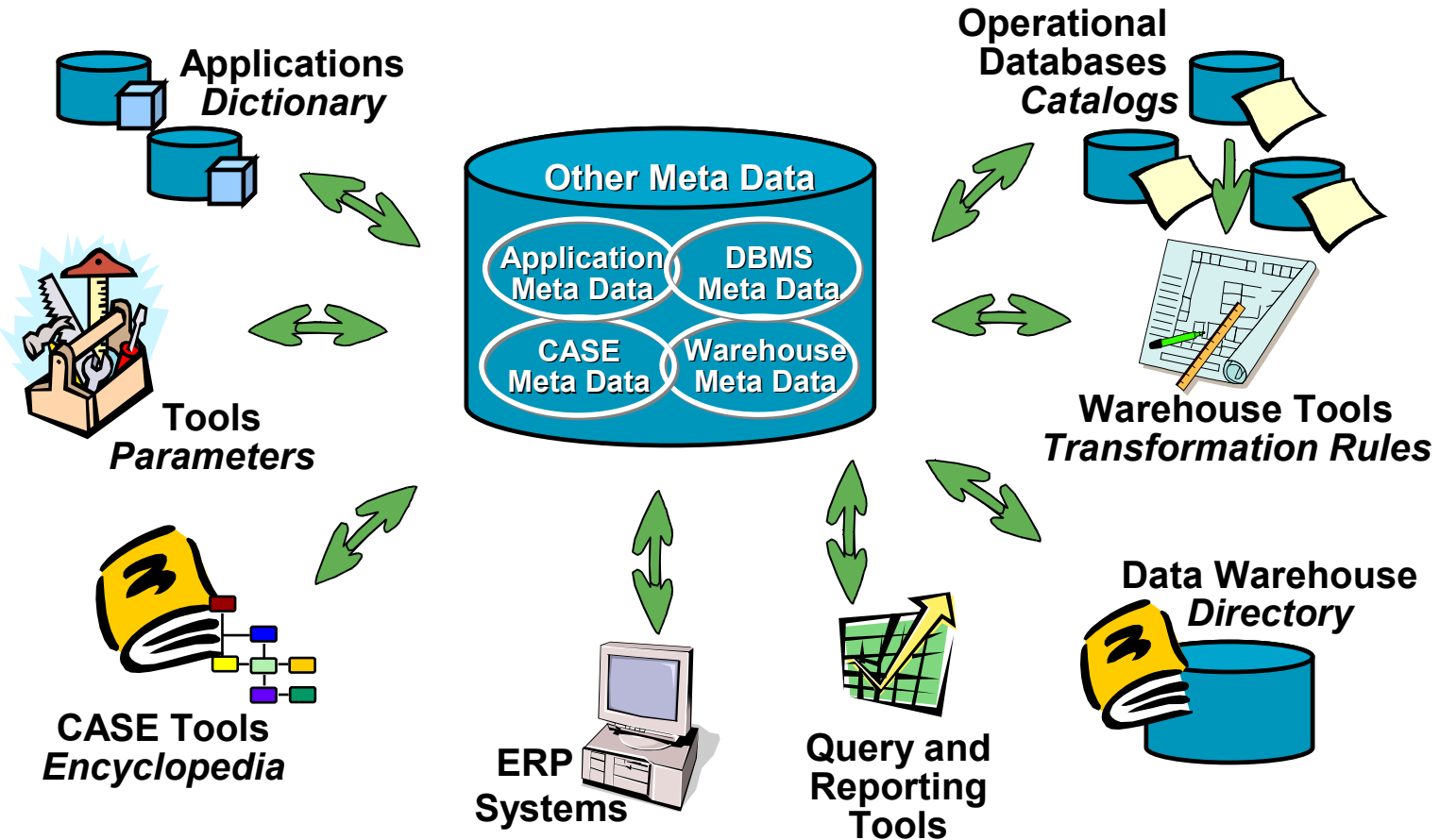
- **Technical metadata is for the technician for work such as maintenance, development, administration, and so forth.**
- **Information about technology**
 - **Examples:**
 - length of a field,
 - shape of a data structure
 - name of a table
 - physical characteristics of a field
 - number of bytes in a table
 - indexes on a table

David Marco – Technical Metadata

- **Meta data that supports a company's technical users and IT staff**
 - **Examples:**
 - The method a technical analyst uses for development and maintenance of the data warehouse
 - Physical table and column names
 - Data mapping and transformation logic
 - Source systems
 - Foreign key and indexes
 - Security
 - ETL process names

Enterprise Metadata Repository

Departmental "Meta Data Marts"



METAWRIGHT

Metadata Integration

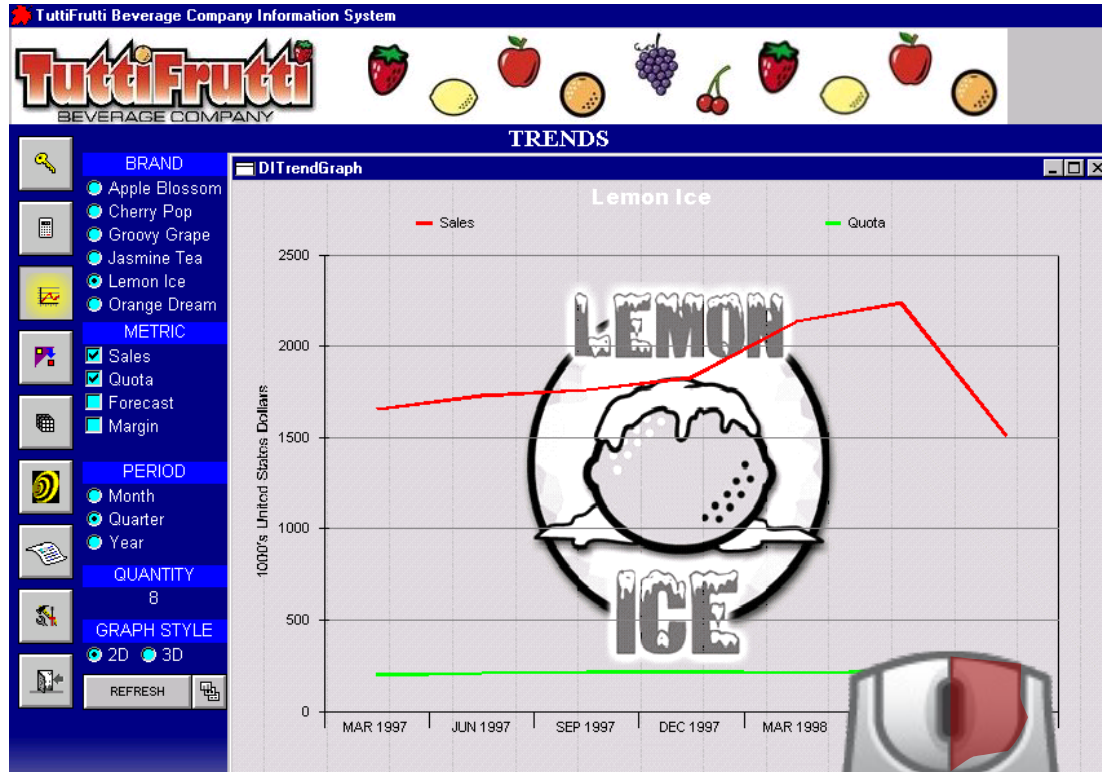


- **Metadata integration across these multiple tools has become the holy grail.**

▪ **But just like the holy grail, it seems to be a never-ending quest fueled by the belief that technology is going to solve all our problems.**

Rick Sherman
DM Review Magazine, January 2006

Business Metadata - Right Mouse Click



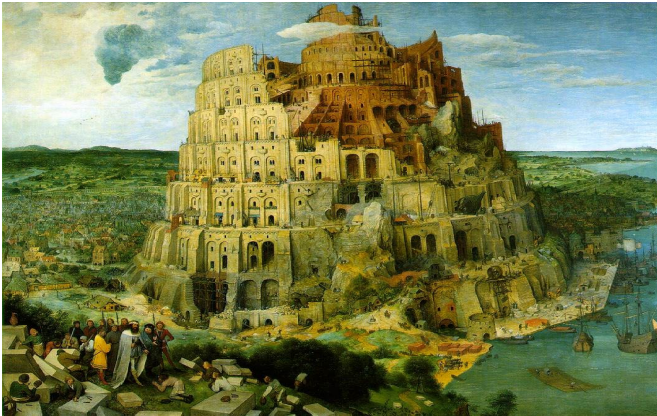
PROD_STAT_CDE:

C - CONCEPT - conceptual stage, no product planning performed, cataloged as idea for future development

METAWRIGHT

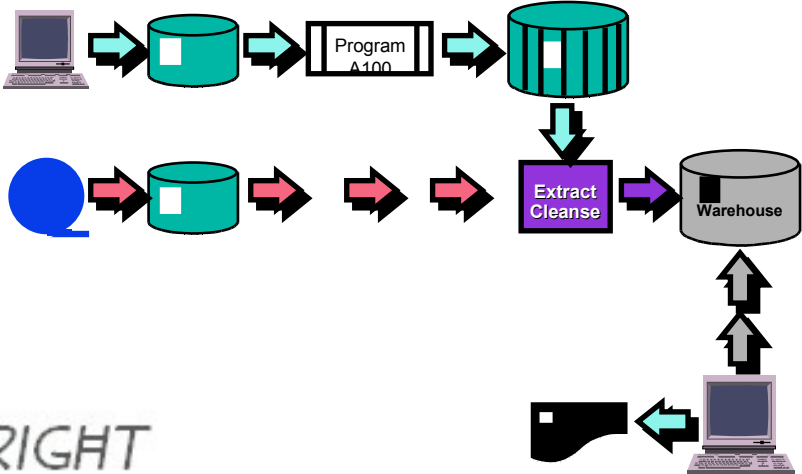
The HUH? Factor

UNDERSTANDIBILITY Synonyms, Homonyms, Artifacts

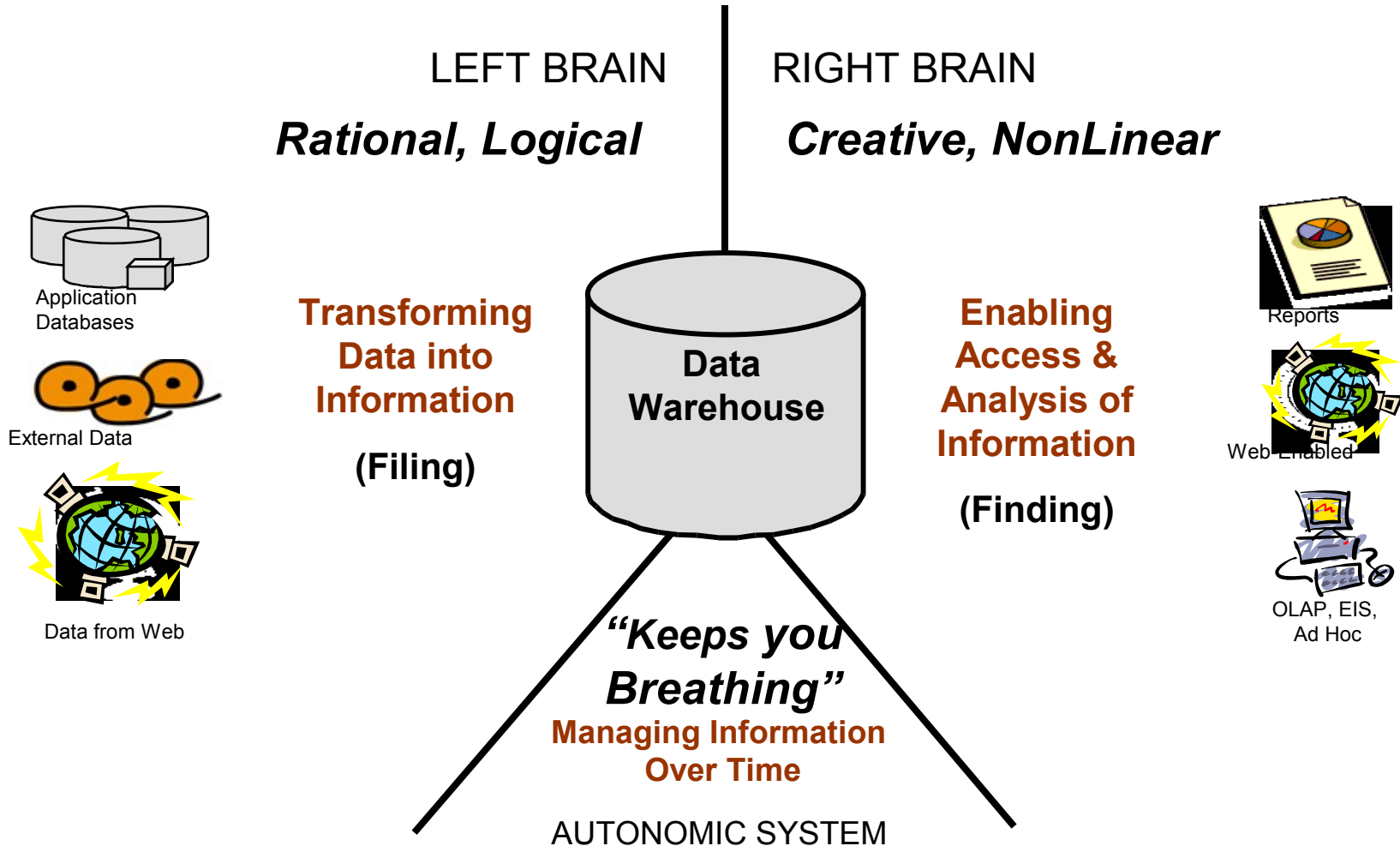


CREDIBILITY

TRACEABILITY



Data Warehouse - Brain



METAWRIGHT

Technical Metadata



- **Provides the “Wiring Diagram” for the DW Environment**
- **Objectives:**
 - Detailed source-to-target mappings
 - Impact Analysis
 - Reusability
 - Auditability
 - “The Defense of the Warehouse”

How Can I Get More?



**Different Data
Same Time Slice**

**Same Data
Different Time Slice**

**Different Data
Different Time Slice**

William H. Inmon – Business Metadata

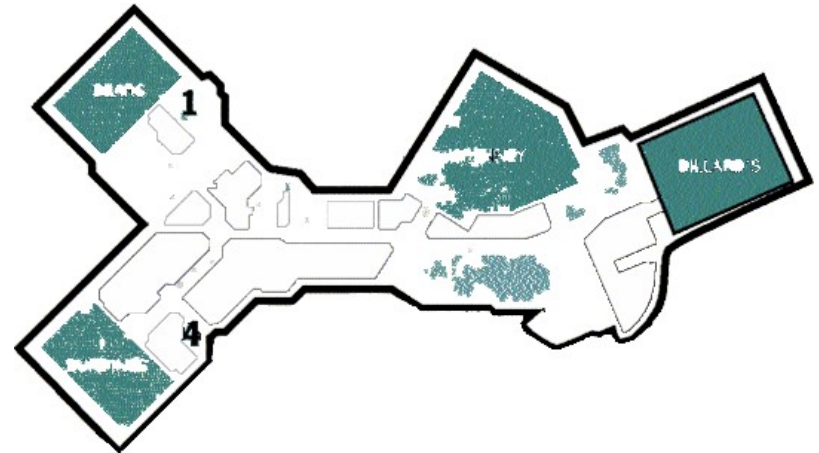
- **Data for the business person.**
- **Business metadata is in the terminology of the business person and is for the business person.**
 - **Examples:**
 - product, expense, revenue, profit, bookings, sales pipeline, sales closed business..

David Marco – Business Metadata

- **Meta data that supports a company's business users**
 - **Examples:**
 - Business terms and definitions for entities and attributes
 - Subject area names
 - Query and report definitions
 - Report mappings
 - Data stewards
 - ETL process names

Business Metadata

- **Serves as the end-user “Mall Map” to the Data Warehouse**
- **Objectives**
 - Context
 - Understandability
 - Searchability
 - Usability
 - Hide technological constraints



Metadata Navigation

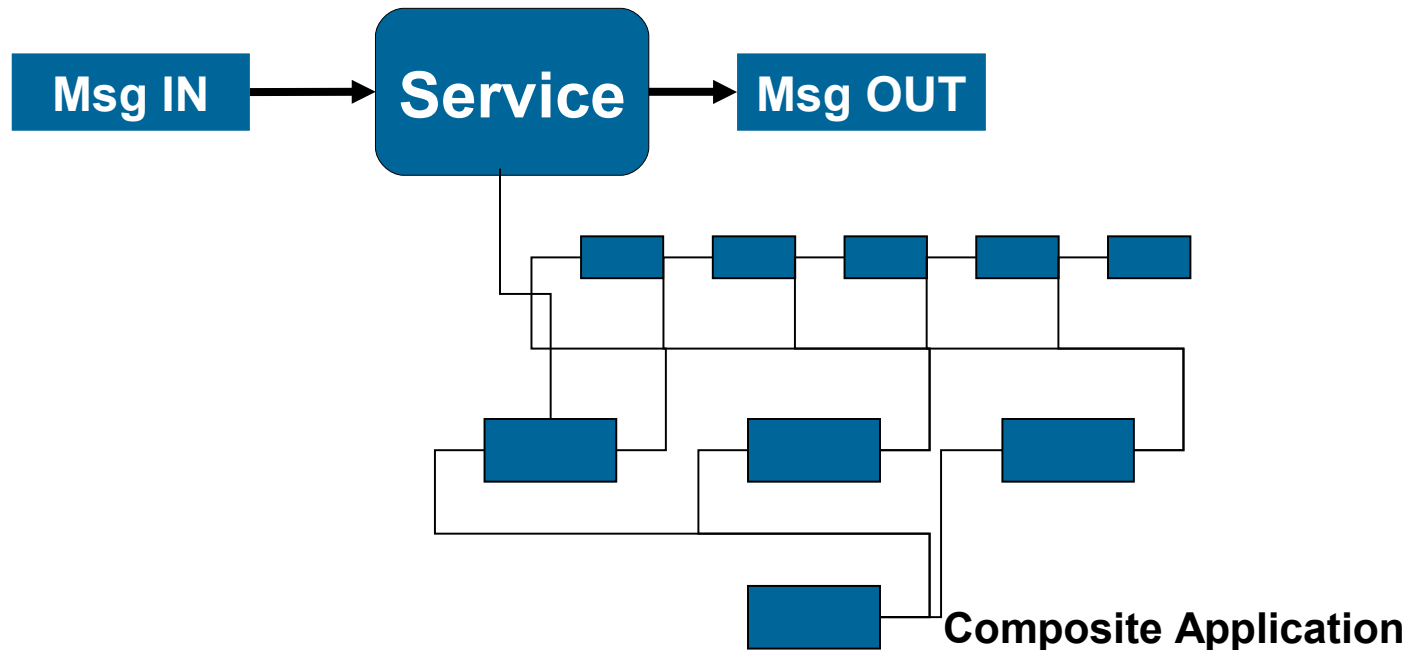
- **What we know**
 - Remember the path
- **What we should know**
 - Find other paths
- **What we don't know**
 - Discover new path

Navigation Tool Requirements

- **Provide business and technical analysts with methods to browse and interpret metadata:**
 - Generate dynamic ‘spider web’ navigation based on object relationships (Mall Map)
 - Dynamic mapping and visualization of relationships between two or more objects selected by an analyst (Object Shopping Cart)
 - Surround reports with relevant metadata including data provenance, definitions, stewardship and related data and metadata
 - Link metadata through set of services from multiple sources and display in a single interface
 - Present Data in Context:
 - Person? Place? Process? Period? Product? Purpose?
 - Pivot view of metadata around any object or report

Key Rationalization for SOA

- **Decrease cost of Integration**
- **Increase speed of delivery of business functionality**



SOA Metadata Standards

- **Simple Object Access Protocol (SOAP)**
 - Describes the runtime call to a service – what data to pass in, which method to call, which security credentials to apply...
- **Web Services Description Language (WSDL)**
 - Describes the interface of the service – which methods are available and which schema each method expects as inputs and outputs
- **Universal Description, Discovery and Integration (UDDI)**
 - Describes all the services available and where they can be found

Metadata Domains

Zachman Framework for Enterprise Architecture

		WHAT	HOW	WHERE	WHO	WHEN	WHY	
Model	CONTEXTUAL Scope	List of things	List of Processes	List of Locations	List of Organization Units	List of Events	List of Business Goals/Stat.	Business
	CONCEPTUAL Business Model	Business Entity Model	Business Process Model	Logistics Network	Work Flow Model	Master Schedule	Business Plan	
Create	LOGICAL System Model	Logical Data Model	Application Process Model	System Network Model	Human Interface Paradigm	Processing Structure	Business Rule Model	Technical
	PHYSICAL Technology Model	Physical Data Model	Application Structure Chart	Network Technology Model	Presentation Architecture	Control Structure	Rule Design	
Assemble	OUT-OF-CONTEXT Components	Data Definition	Program	Network Components	Interface Components	Timina Definition	Rule Specification	
	PRODUCT Functioning System	DATABASE	APPLICATION	NETWORK	ORGANIZATION	SCHEDULE	STRATEGY	Operational

Data about Data

- **Business Metadata**
 - Semantic definitions and business context for enterprise information systems.
- **Technical Metadata**
 - Logical and physical design and blueprint information for enterprise information assets.
- **Operational Metadata**
 - Information describing the activity of information systems in production.

Metadata Integration

Data Rationalization

*META*WRIGHT

Enterprise Resource Planning

- **Other Peoples Models!**

Oracle
Financials

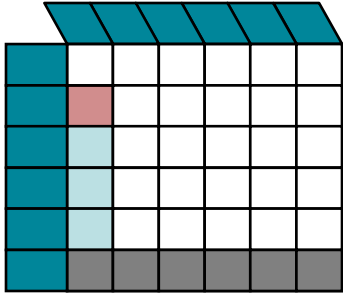


PEOPLE®
Soft
METAWRIGHT

Tracking Data Transformation

CONCEPTUAL Business Model	Business Entity Model
LOGICAL System Model	Logical Data Model
PHYSICAL Technology Model	Physical Data Model
OUT-OF-CONTEXT Components	Data Definition

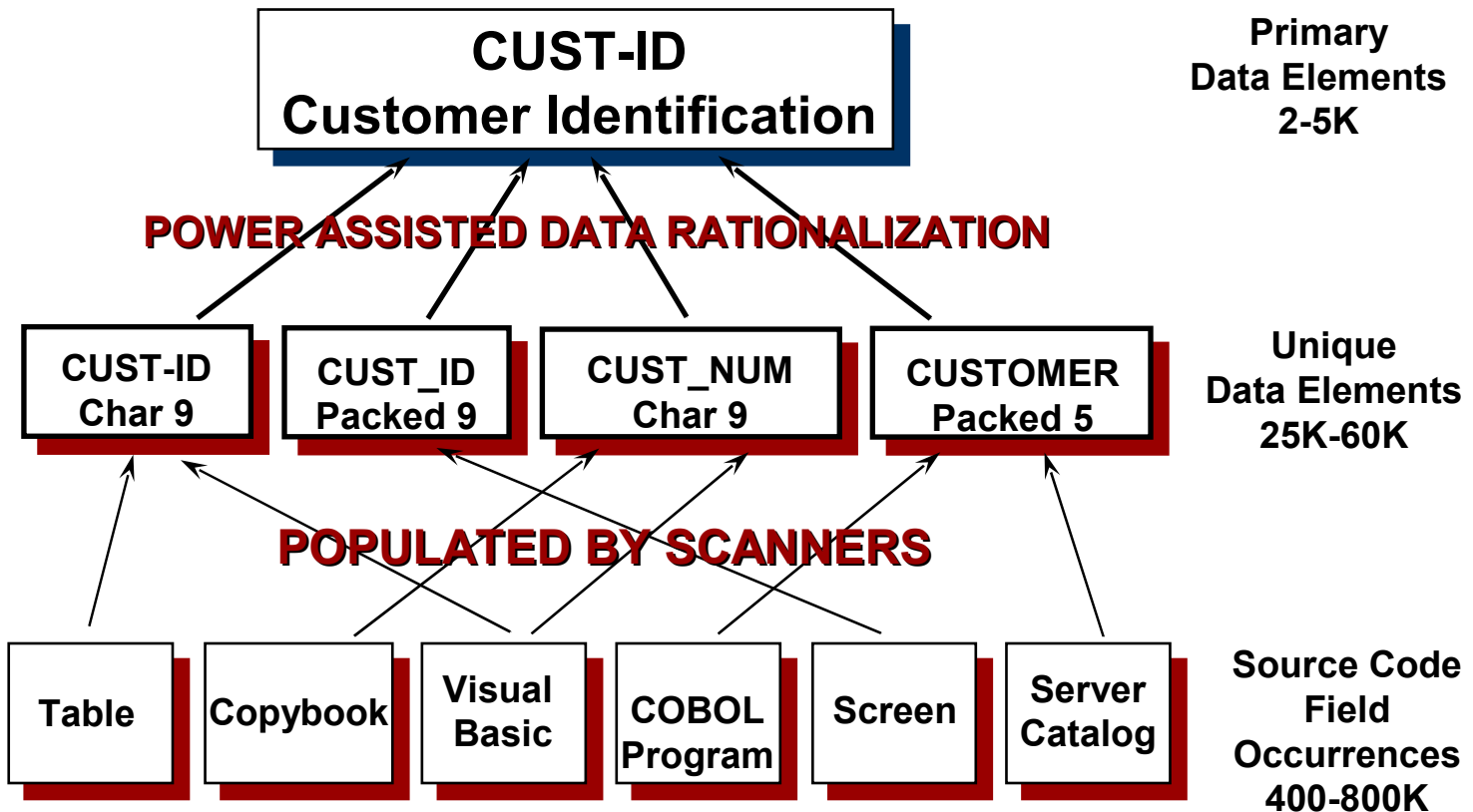
Subject Areas, **Business Data Element** 1
 :
Entities, **Attributes**, Relationships M
 :
Tables, **Columns**, Indexes, Copybooks M



Enterprise Framework Standards

- **Short Name:**
 - Inventory Definition
- **Formal Name:**
 - Enterprise Inventory Semantic Model
- **Description:**
 - This semantic model is the enterprise inventory definition, and is the business perspective contributed by and for the executive leaders.
- **Formulation:**
 - The model defines the set of inventory to answer the question "What?". The answer is expressed by a contiguous inventory model which relates one 'business entity' through a 'business relationship' to another peer 'business entity'.
- **Components:**
 - Business Entity
Business Relationship

Metadata Integration



Repository Path Report

Element Columns

Business Data Element: **Customer Identification Number** *Numeric code which uniquely identifies a customer.*

Relational Column: **cust_id**

Table Has Column: **CUSTOMER.cust_id**

Relational Column: **CUST_ID**

Table Has Column: **CUSTOMER.CUST_ID**

SQL Statement: **customer_listing.sql**

Relational Column: **Cust_Nbr**

Table Has Column: **CUST.Cust_Nbr**

Table Has Column: **ITEM.Cust_Nbr**

Table Has Column: **ORD.Cust_Nbr**

Data Rationalization Process

- **Naming Standards - Order**
 - Class Words
 - Prime Words
 - Modifying Words
- **Determine Scope**
- **Create Team - Collect Supporting Data**
- **Determine Business Meaning (Local/Global)**
- **Identify/Create Business Entities**
- **Create Aliases**

Naming Standards

- **Class Words**

- High-level type of element
- Relate to the data type of the element.
 - AMT Amount
 - CD Code
 - DT Date

- **Prime Words**

- General Topic or Subject Area
 - CUSTOMER, INVOICE, PRODUCT

- **Modifying Words**

- Provides further detail to the element name.
 - DY Day
 - TOT Total
 - FNL Final

- **Order**

- 1 Prime Word + *n* modifying words + 1 Class Word
- CUSTOMER + DY + TOT + AMT

Scope – Supporting Data - Team

- **Project Scope**

- Application/Project related set of files/tables
- Topic or Subject Area

- **Supporting Data Collection**

- Repository query showing data elements, entities and attributes. Report of previously defined glossary items.
- Documentation sources.

- **Team Composition**

- Business subject matter expert – data steward
- Data Base expert
- Data analyst/modeler
- Trained Facilitator

Business Meaning

- **Develop list of elements collected into groupings by topic (Class Word)**
- **Break down topic groupings in subsets of like looking/sounding elements**
 - **CUSTOMER IDENTIFIER**
 - Cust_ID
 - Cust_Num
 - Order_Cust_Num
 - Customer_Name
- **Confirm/Document business definition**

Data Definitions



- **The Sphinx: “To learn my teachings, I must first teach you how to learn.”**
- **Tautology, defined by itself:**
 - **Cust_ID** is the ID of the Customer
- **Not too big, Not too small**

Business Metadata: How to Write Definitions, B. O’Neil – b-eye Network - 2005

*META*WRIGHT

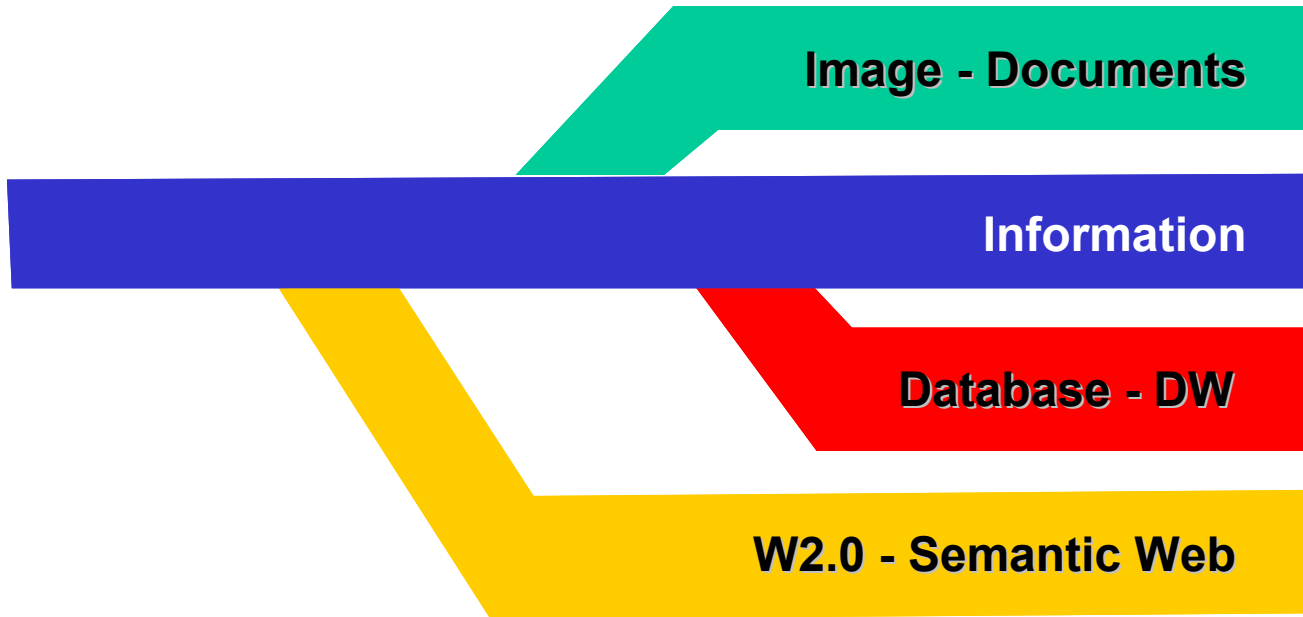
Primary Elements - Aliases

- **Define Standard Business Name, Data Type and Description**
- **Potential to add Dublin Core Metadata Attributes**
- **Business Name can be a union of several elements.**
 - CUSTOMER-NAME =
CUSTOMER-FIRST-NAME + CUSTOMER-LAST-NAME
- **Identify “Enterprise” Business Names (Qualifier)**
- **Develop set of relationships to business, logical or physical aliases.**
 - Same business meaning, different name
 - Same business meaning, technology constrained names

Interoperability

*META*WRIGHT

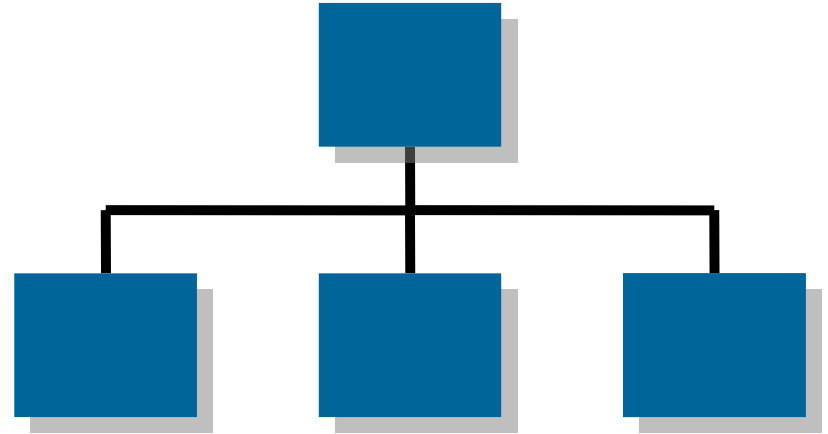
Grand Unified Theory



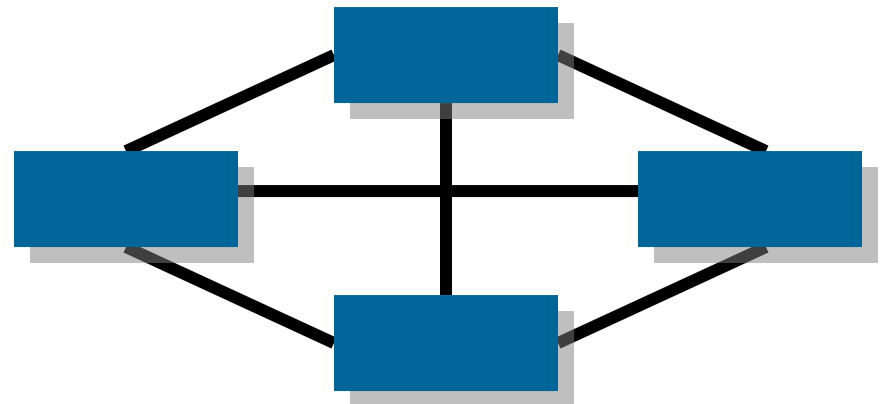
METAWRIGHT

Database Models

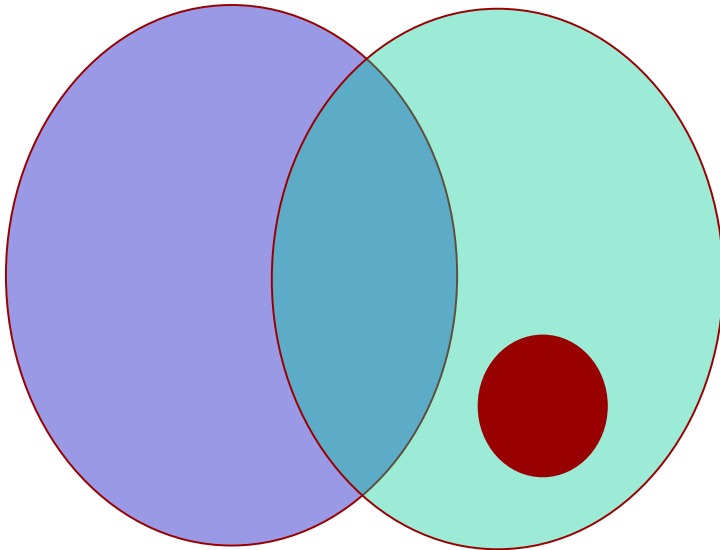
- **Hierarchy model**



- **Subject Relationship model**



Relational Model - DataSET



- **Unique Key Identifiers**
- **Eliminate M:M relationships**
- **Inside / Outside Joins**
- **Access via SQL**

The 12 (no, 13) Rules of Relational Theory

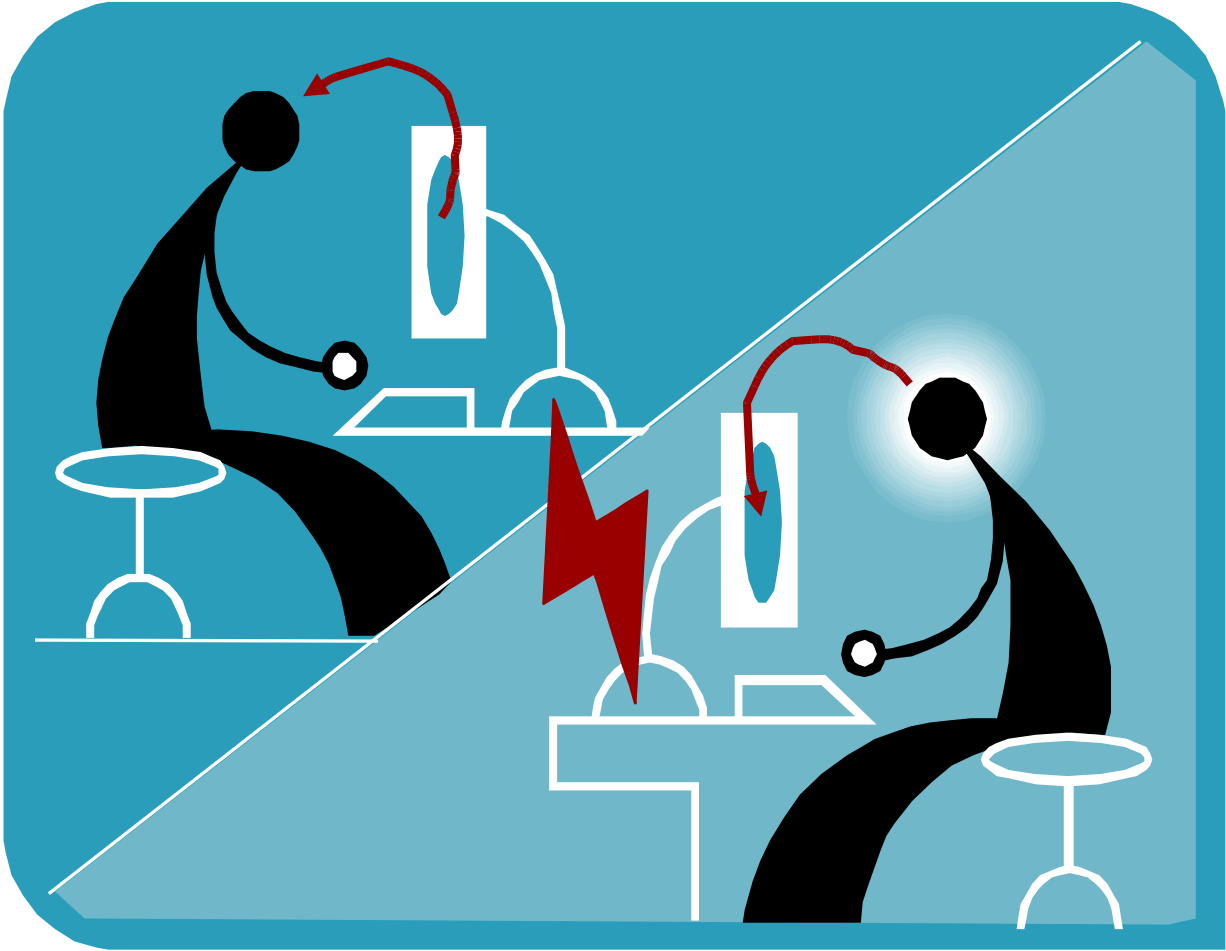
"A Relational Model of Data for Large Shared Data Banks" E.F.Codd,1970

Enterprise Metadata Integration

Zachman Framework for Enterprise Architecture

	WHAT	HOW	WHERE	WHO	WHEN	WHY
CONTEXTUAL Scope	List of things	List of Processes	List of Locations	List of Organization Units	List of Events	List of Business Goals/Stat.
CONCEPTUAL Business Model	Business Entity Model	Business Process Model	Logistics Network	Work Flow Model	Master Schedule	Business Plan
LOGICAL System Model	Logical Data Model	Information Process Model	System Network Model	Human Interface Paradigm	Processing Structure	Business Rule Model
PHYSICAL Technology Model	Physical Data Model	Information Structure	Network Technology Model	Presentation Architecture	Control Structure	Rule Design
OUT-OF-CONTEXT Components	Data Definition	Program	Network Components	Interface Components	Timing Definition	Rule Specification
PRODUCT Functioning System		OPERATION	NETWORK	ORGANIZATION	SCHEDULE	STRATEGY

Semantic Web



METAWRIGHT

Tim Berners-Lee – Vision

- **“I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers. A ‘Semantic Web’, which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines.”**

Semantic Web

- **At its core, the semantic web comprises a philosophy, a set of design principles, collaborative working groups, and a variety of enabling technologies.**
- **Some of these include**
 - Resource Description Framework (RDF),
 - Data interchange formats
 - Notations such as RDF Schema (RDFS)
 - Web Ontology Language (OWL)

Wikipedia.com

Semantic Relationships

- **Element Definition**
- **Vocabularies (Glossary)**
- **Taxonomies (Controlled Vocabularies)**
- **Thesauri**
- **Ontology**

Dublin Core

- **1994 – Meeting of the minds of Liberians and IT people in Dublin, OH**
- **1995 – Dublin Core - Library card for Web**

Identifier

Title

Description

Subject

Date

Relation

Creator

Publisher

Contributor

Format

Coverage

Type

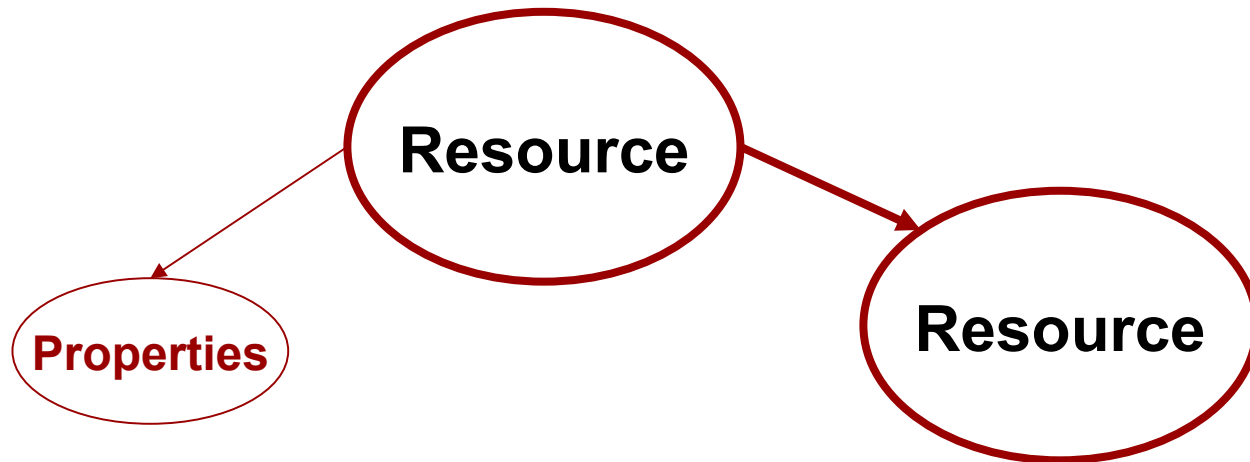
Source

Rights

Language

Metadata Modeling – DCMI 2003-2005

- **Similar to Resource Description Framework (RDF) – Semantic Web**
- **Vocabulary for describing metadata**
 - Resource with Properties
 - Resources relate to other resources



Dublin Core Refinements

Abstract	Is referenced by
Access rights	Is replaced by
Alternative	Is required by
Audience	Issued
Available	Is version of
Bibliographic citation	License
Conforms to	Mediator
Created	Medium
Date accepted	Modified
Date copyrighted	Provenance
Date submitted	References
Education level	Replaces
Extent	Requires
Has format	Rights holder
Has part	Spatial
Has version	Table of contents
Is format of	Temporal
Is part of	Valid

ISO Data Element

- **Should be registered according to the Registration guidelines (11179-6)**
- **Will be uniquely identified within the register (11179-5)**
- **Should be named according to Naming and Identification Principles (11179-5)**
- **Should be defined by the Formulation of Data Definitions rules (11179-4)**
- **May be classified in a Classification Scheme (11179-2)**

ISO 11179 – Metadata Registry Standard

- **Part 1 - Framework**
- **Part 2 - Classification**
- **Part 3 - Registry metamodel and basic attributes**
- **Part 4 - Formulation of data definitions**
- **Part 5 - Naming and identification principles**
- **Part 6 - Registration**

Ontology - Taxonomy

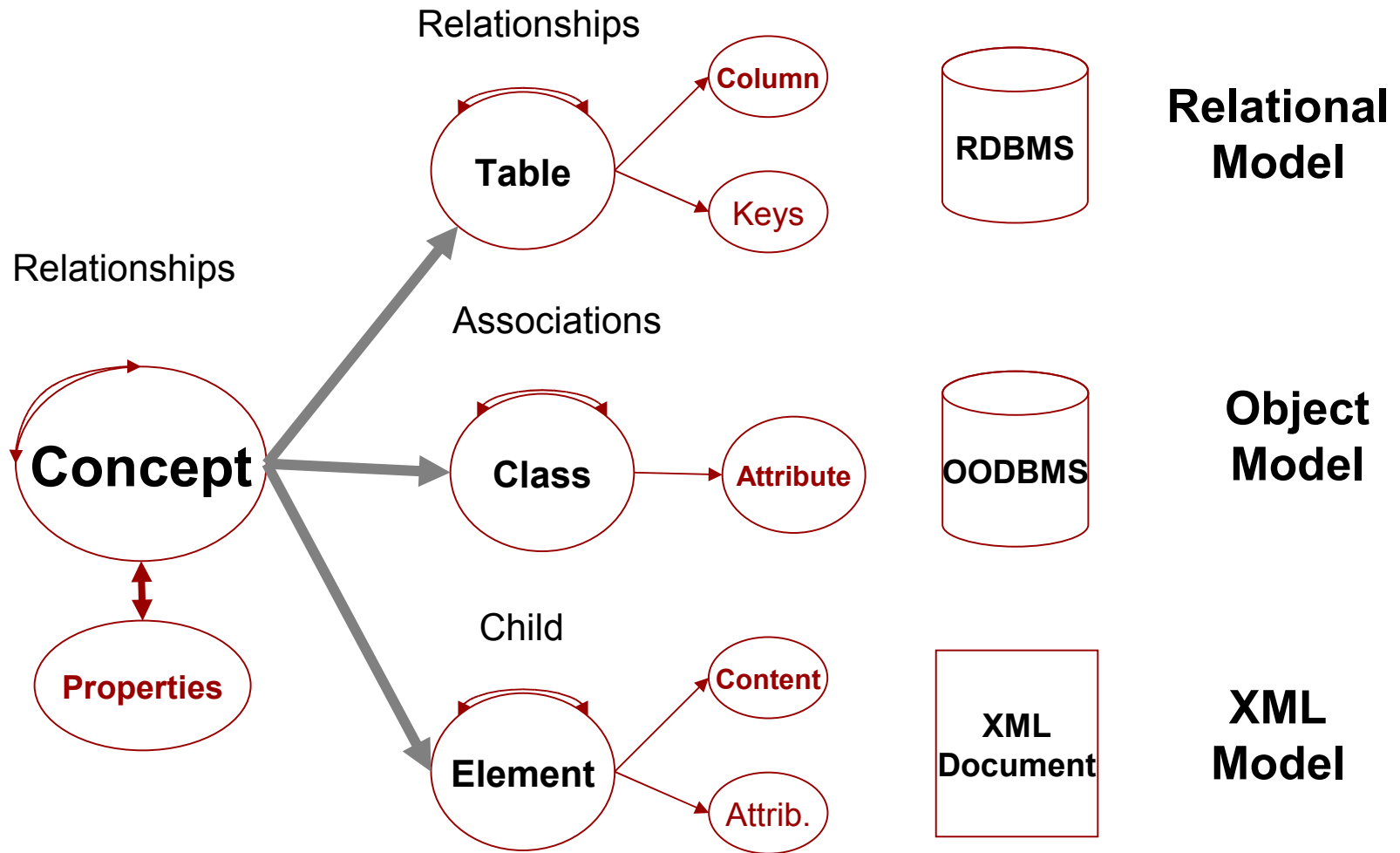
- **Ontology**

- Definition of terms used to describe an area of knowledge within a specific domain or subject area.
- Encapsulate knowledge to share with other domains

- **Taxonomy**

- Classification scheme, typically described in a hierarchical structure.
- Supports drill down and multiple paths.

MetaModel Integration



METAWRIGHT

Metadata Stewardship

- **“Governance” Steward**
 - Responsible for the decision about **WHAT** metadata the enterprise will collect and maintain the metadata represented by the **Column**
 - **REPRESENTS DATA CREATORS**
- **“Process” Steward**
 - Responsible for the decision about **HOW** the enterprise will collect and maintain the metadata represented by the **Row**
 - **REPRESENTS DATA CONSUMERS**

Metadata Management 2.0?

- **Metadata Lifecycle - BTO**
- **Master Data (CDI) Integration**
- **Integration of Non-textual Metadata**
- **Cross pollination of Metadata Standards**
 - **Semantic Web**
 - Resource Description Framework (W3C)
 - Dublin Core (Non W3C)
 - **ISO/IEC 11179**
- **IT Information Warehouse Products**

Donald J. Soulsby
DSoulsby@MetaWright.com

*META*WRIGHT