



Creating the Golden Record

Better Data through Chemistry

Donald J. Soulsby
metaWright.com

META^WRIGHT

Agenda

- **The Golden Record**
- **Master Data**
 - Discovery
 - Integration
 - Quality
- **Master Data Strategy**

DAMA – LinkedIn Group

C. Lwanga Yonke

- Information Quality Practitioner

...Spewak advocated using data dependency to determine the ideal sequence in which applications should be developed and implemented: “Develop the applications that create data before those that need to use that data” (p.10).

Architecture Advocates

William Smith

- Entity Lifecycle

Clive Finkelstein

- Information Engineering – CRUD

Ron Ross

- Resource Life Cycle Analysis

CRUD in a Perfect World

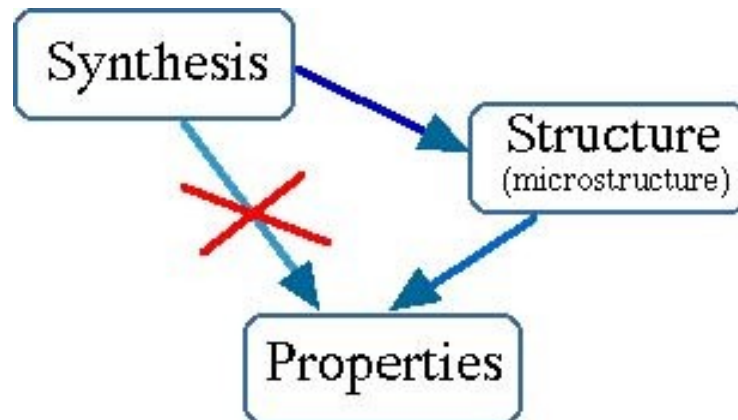
Entity Process	Customer	Customer Order	Customer Account	Customer Invoice	Vendor Invoice	Product
Receive Customer Order	R	C	CR			
Process Customer Order	CRU		RU			R
Maintain Customer Order	U		U		RU	
Terminate Customer Account	U		U		RU	
Fill Customer Order	RU		RU			RU
Ship Customer Order			U		C	
Validate Vendor Invoices					R	
Pay Vendor Invoices					RU	
Invoice Customer	RU		RU	C		
Maintain Inventory						CRUD

Canonical Synthesis

Broadly speaking, materials scientists investigate two types of phenomena. Both are based on the microstructures of materials:

...

ii. How do these microstructures influence the properties of the material (such as strength, electrical conductivity, or high frequency electromagnetic absorption)?



<http://www.its.caltech.edu/~matsci/WhatIs2.html>

METAWRIGHT

Business VS Development Life Cycles

Zachman Framework for Enterprise Architecture

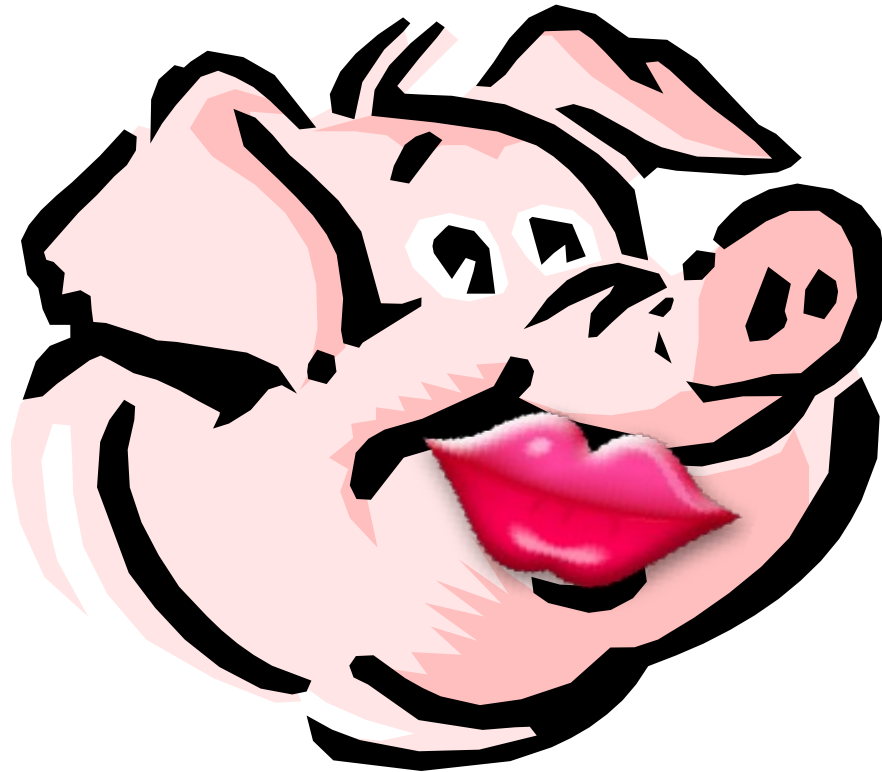
	WHAT	HOW	WHERE	WHO	WHEN	WHY
CONTEXTUAL Scope	List of things	List of Processes	List of Locations	List of Organization Units	List of Events	List of Business Goals/Stat.
CONCEPTUAL Business Model	Business Entity Model	Business Process Model	Logistics Network	Work Flow Model	Master Schedule	Business Plan
LOGICAL System Model	Logical Data Model	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
OUT-OF-CONTEXT Components	Data Definition	Program	Network Components	Interface Components	Timing Definition	Rule Specification
PRODUCT Functioning System	DATABASE	APPLICATION	NETWORK	ORGANIZATION	SCHEDULE	STRATEGY

Tibetan Proverb

水头 不清水 尾 浑

“ If upstream is dirty,
downstream will be muddy”

Data Warehousing - ETL



METAWRIGHT

Transmutation



METAWRIGHT

The Secret

Bernard Trevisan, a 15th century alchemist, spent much of his life and a sizable fortune in search of the secret of turning base metals into gold, realized while dying was...

“To make gold, one must start with gold.”

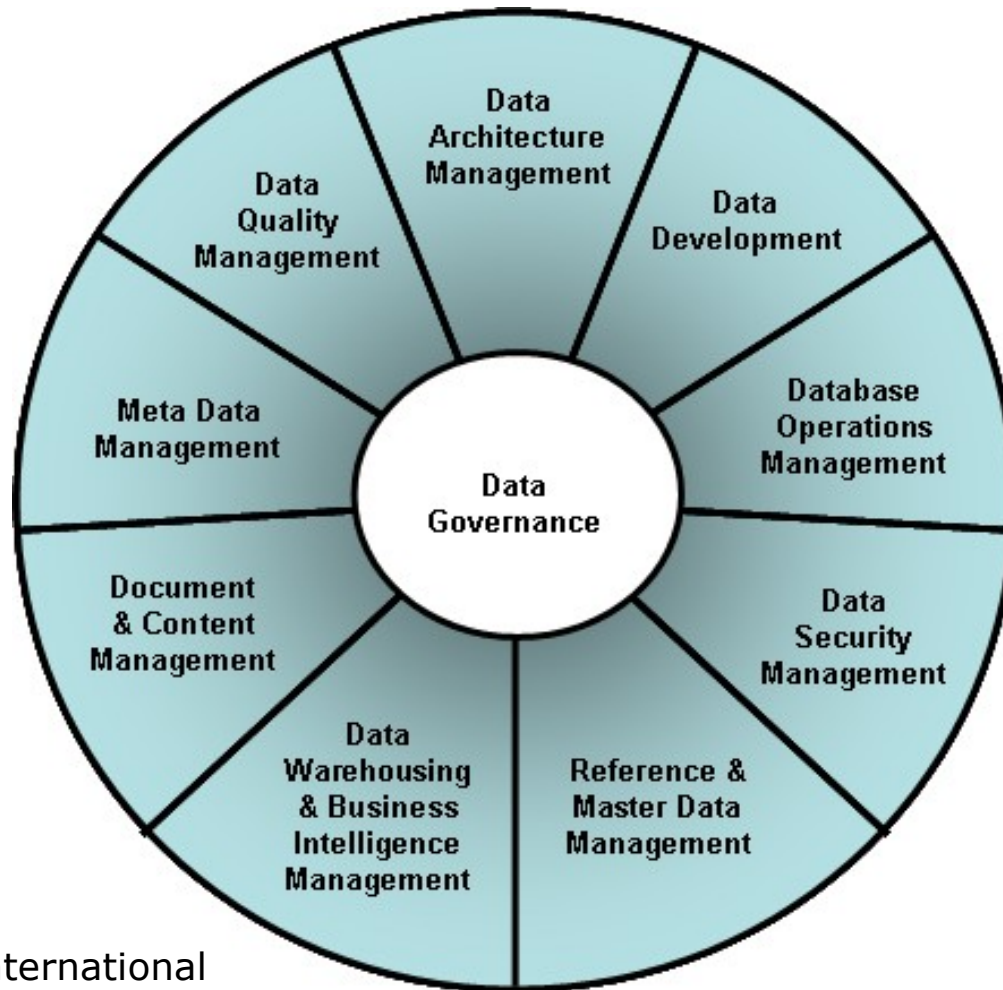
DAMA Guide to the Data Management Body of Knowledge



Reference and Master
Data Management:
Managing golden
versions and replicas

METAWRIGHT

Data Management Body of Knowledge



© 2008 DAMA International

METAWRIGHT

DAMA Dictionary

Reference Data

Any data used to categorize other data, or for relating data to information beyond the boundaries of the enterprise. See master data.

Master Data

Synonymous with reference data. The data that provides the context for transaction data. It includes the details (definitions and identifiers) of internal and external objects involved in business transactions. Includes data about customers, products, employees, vendors, and controlled domains (code values).

DAMA Dictionary

Master Data Management (MDM)

Processes that ensure that reference data is kept up to date and coordinated across an enterprise. The organization, management and distribution of corporately adjudicated data with widespread use in the organization.

Reference & Master Data Management

Ensuring consistency with a “golden version” of data values. One of nine data management functions identified in the DAMA-DMBOK Functional Framework.

Top 13 MDM buzzwords

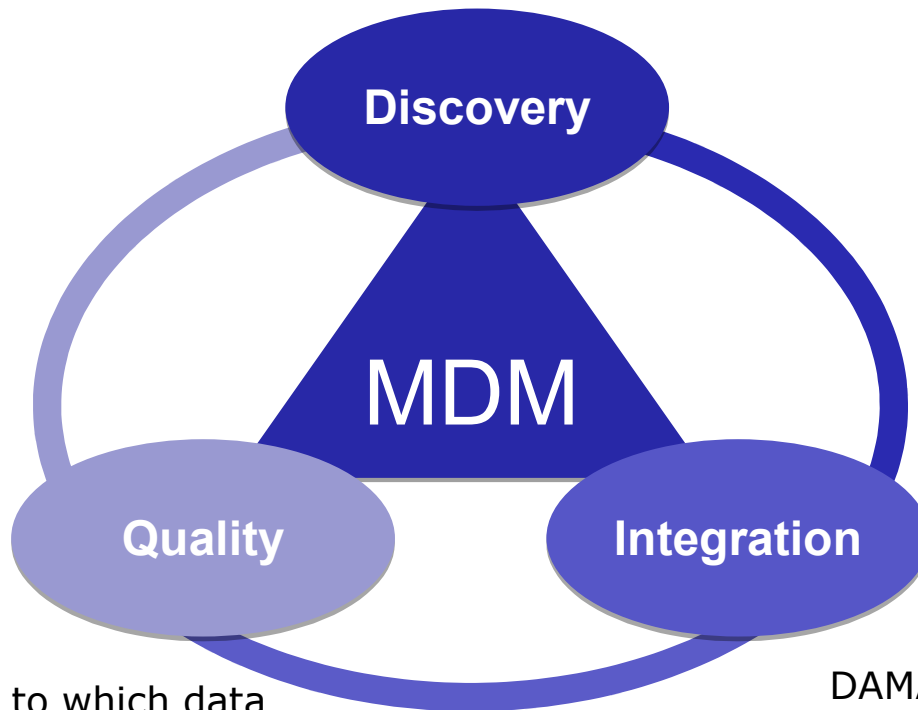
1. Metadata (**Don't say metadata**)

2. Product information management (PIM)
3. **Enterprise master patient index (EMPI)**
4. Data governance
5. Customer data integration (CDI)
6. MDM hub
7. MDM architecture
8. "Collaborative" vs. "analytical" MDM
9. MDM return on investment (ROI)
10. MDM stakeholders
11. Enterprise hierarchy management
12. MDM metrics
13. **Data competency center**

The Quest

Bloor Research: the discovery of relationships
between data elements, regardless of where the
data is stored

dDIQ



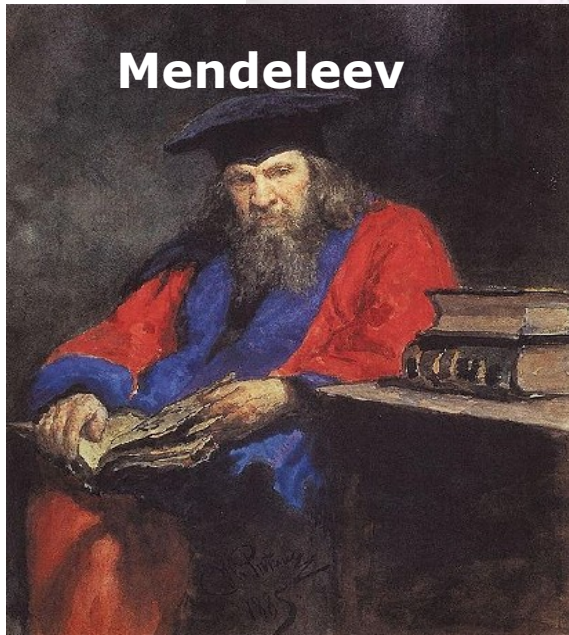
DAMA: The degree to which data
is accurate, complete, timely,
consistent with all requirements
and business rules, and relevant
for a given use.

DAMA: The planned and
controlled transformation and
flow of data across databases,
for operational and/or analytical
use.

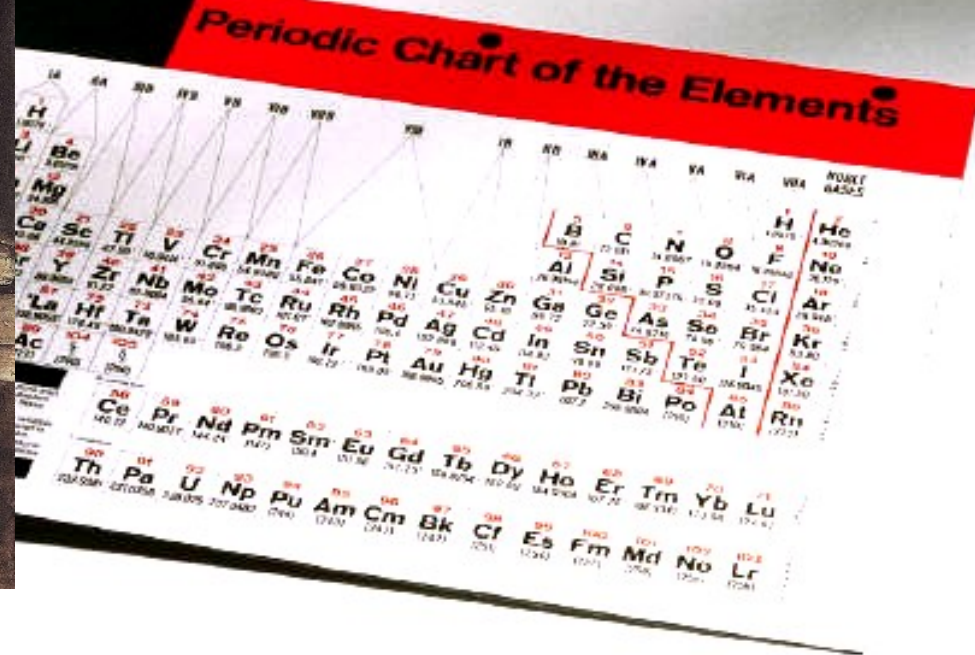
METAWRIGHT

Atomic Data of the Enterprise

Provide and maintain a consistent view of an enterprise's core business information assets



Mendeleev

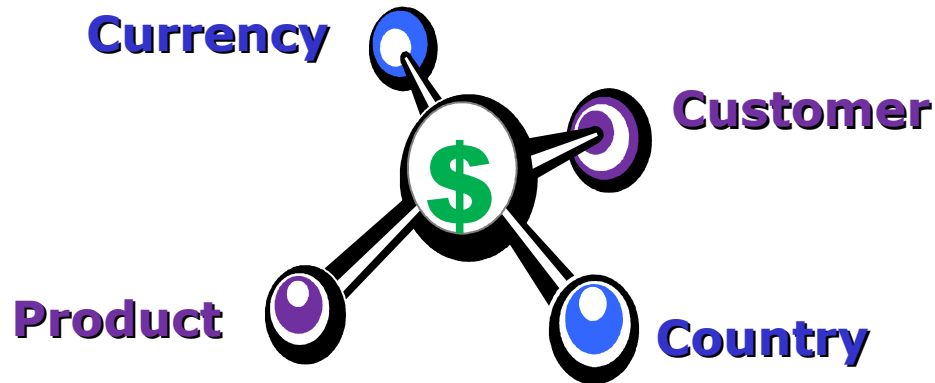
A modern periodic table of elements, titled "Periodic Chart of the Elements". It shows the standard layout of elements with their atomic numbers, symbols, and names. The table is color-coded with various shades of blue, green, and yellow.

ALCHEMY VS. CHEMISTRY

METAWRIGHT

What is Master Data?

Point In Time

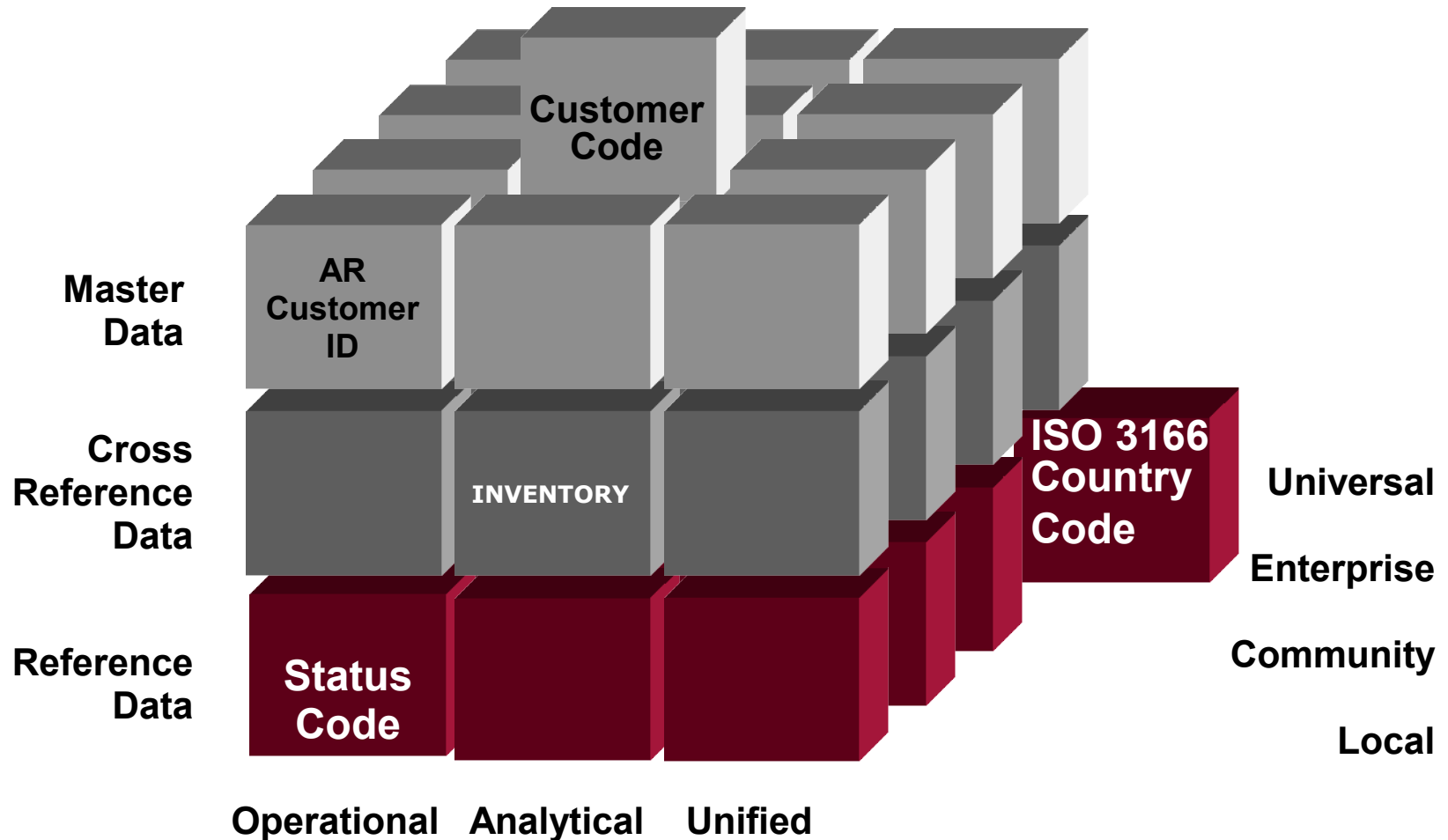


Transaction

A Transaction is relationship of
Master Data + **Reference Files**
+ time stamp + volumes + secondary descriptors

METAWRIGHT

Master Data - Classification



Master Data Types

Operational Master Data

Definition, creation, and synchronization of Master Data required for transactional systems and delivered via service-oriented architecture (SOA).

Analytical Master Data

Definition, creation, and integration, including multiple historical versions, of Master Data required for Enterprise Reporting or Data Warehousing applications.

Unified Master Data

Master Data that is defined and created to apply to both Operational and Analytical applications. Implementation of the “single view of the truth”. It requires achieving agreement on a complex topic among a group of people

Master Data Integration

Universal

Master Data that is created and maintained by an external organization. For reference data it is often developed by a standards organization such as the ISO. Master Data Models are often available to members of an industry trade or professional association

Enterprise

Master Data that represents the common business information assets that need to be agreed on and shared throughout the enterprise

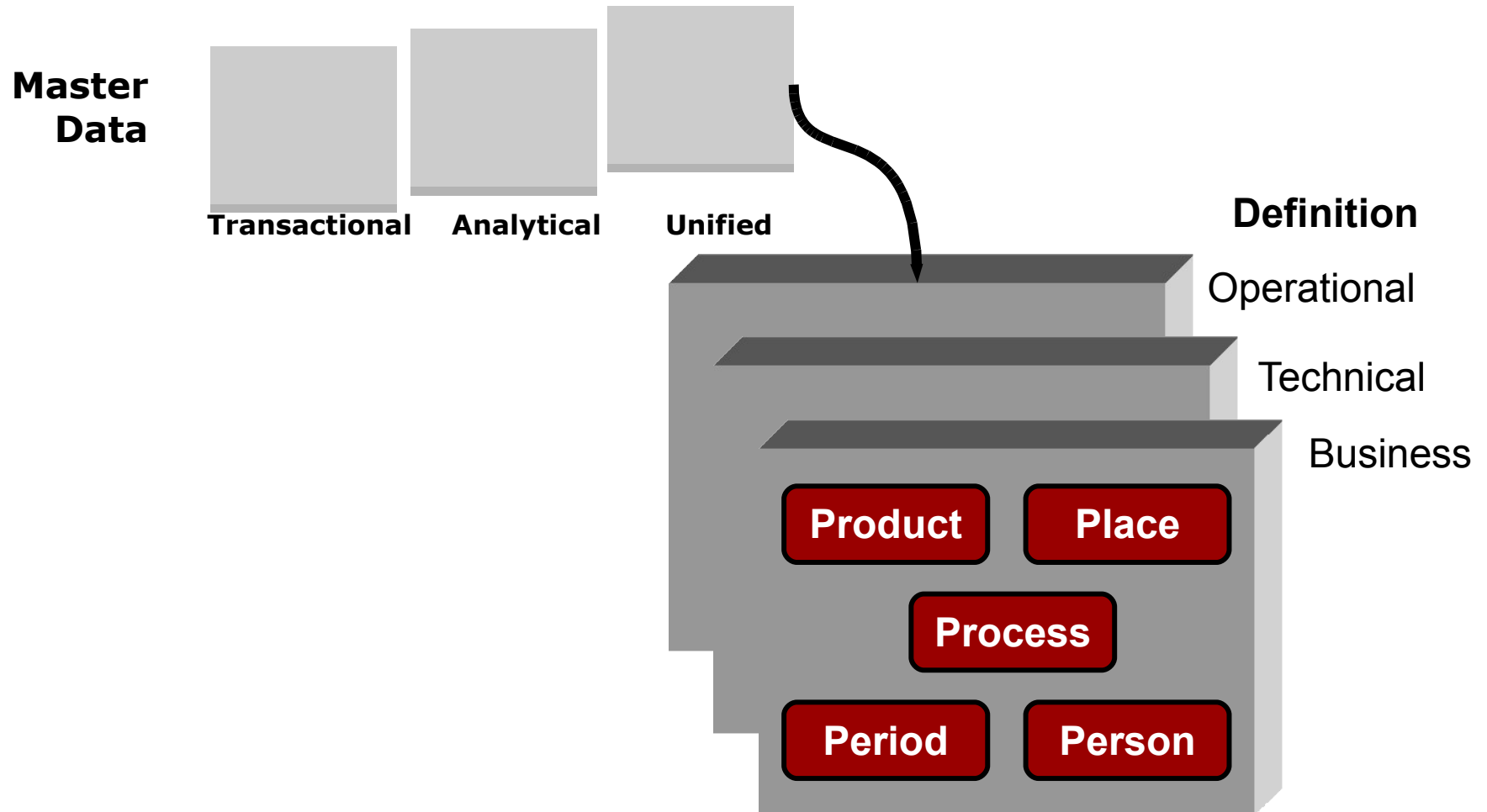
Community

Master Data that is shared between two or more applications. Typically data is replicated from the application that contains the system of record for the Master Data, by federation (key linkage) or propagation (materialized views).

Local

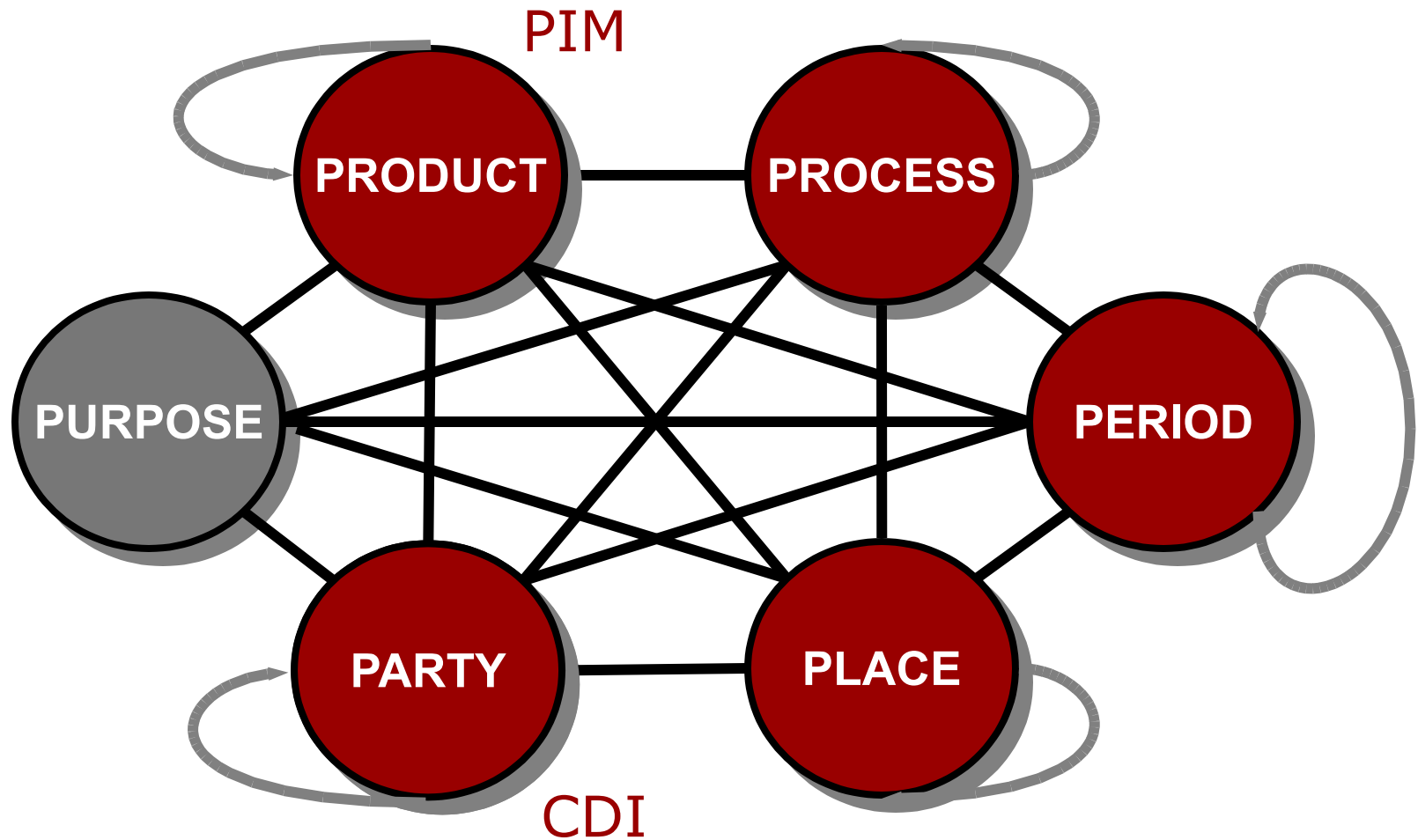
Master or Reference Data that is created and maintained for a single operational application. Much of the Reference Data will remain local, such as Status Codes, for the transactions within the application.

Master Data - Content



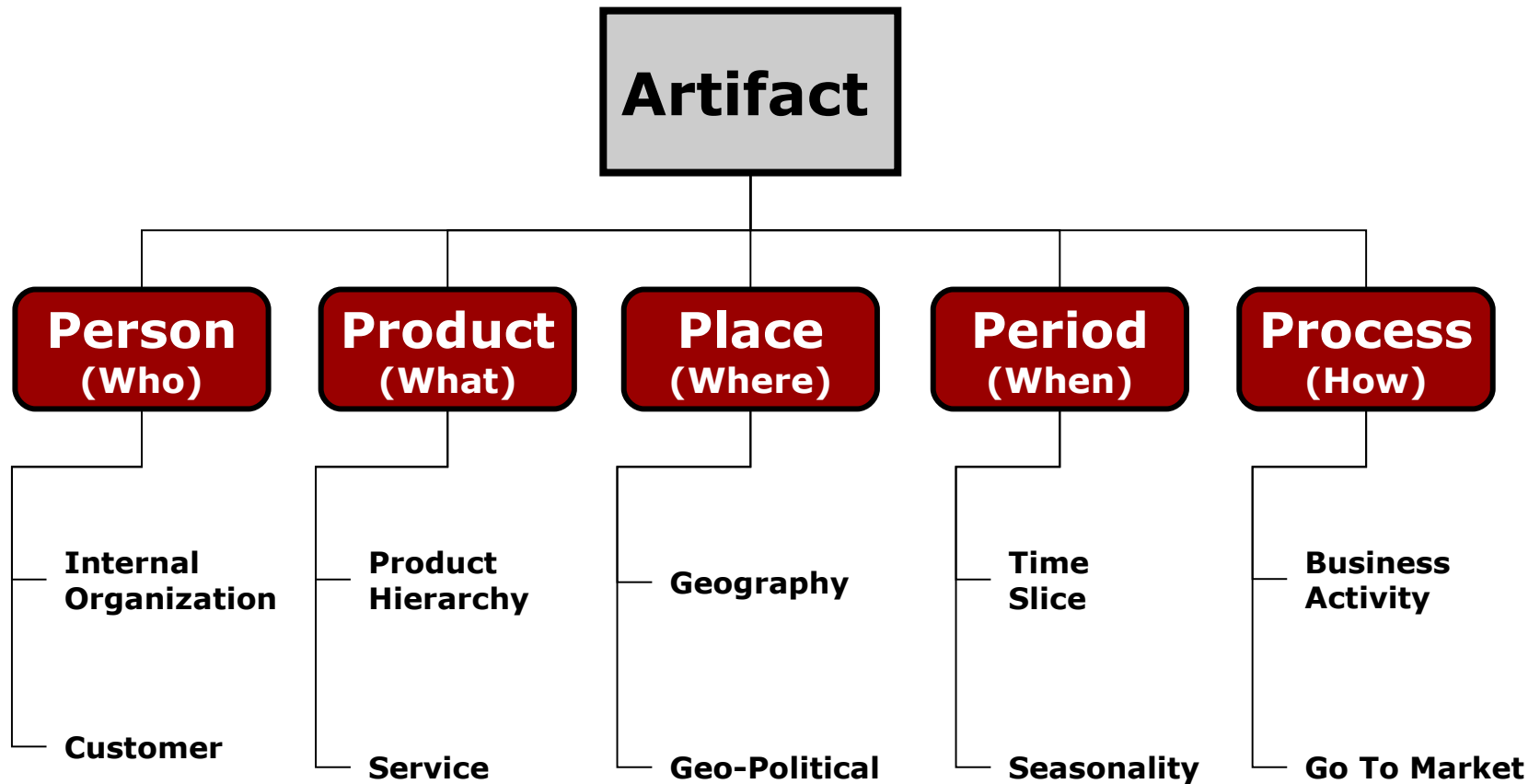
METAWRIGHT

Atomic Primitives



META Wright

Master Data Classification



MDM – Integration Techniques

Data Propagation (Today)

Copies Master Data from one location to another (materialized view).

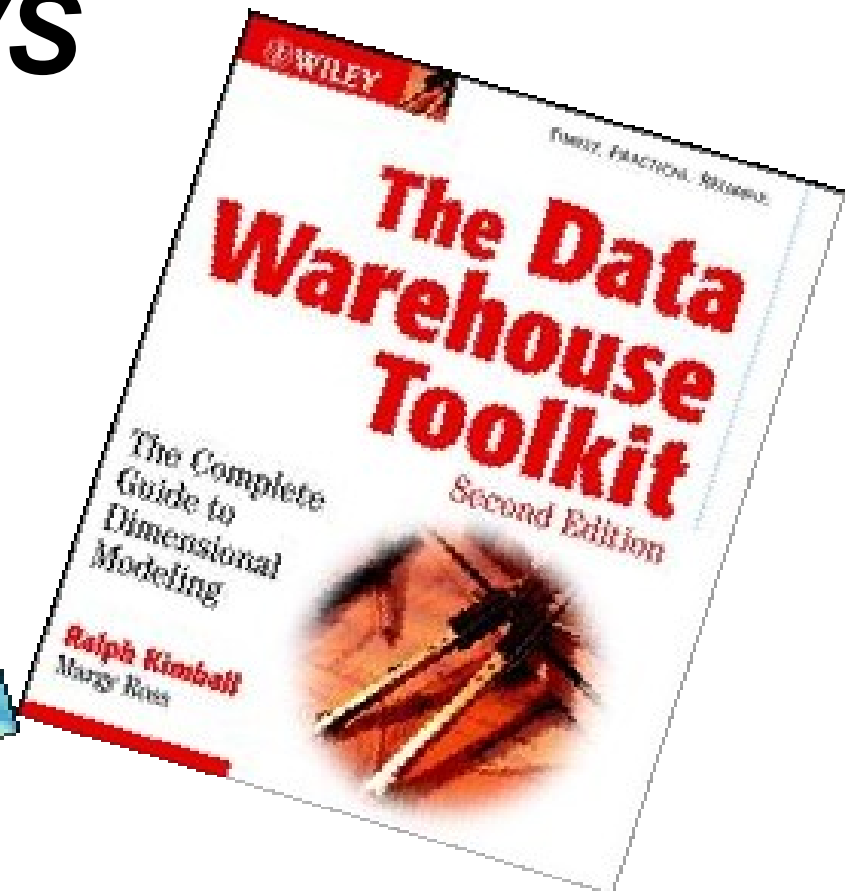
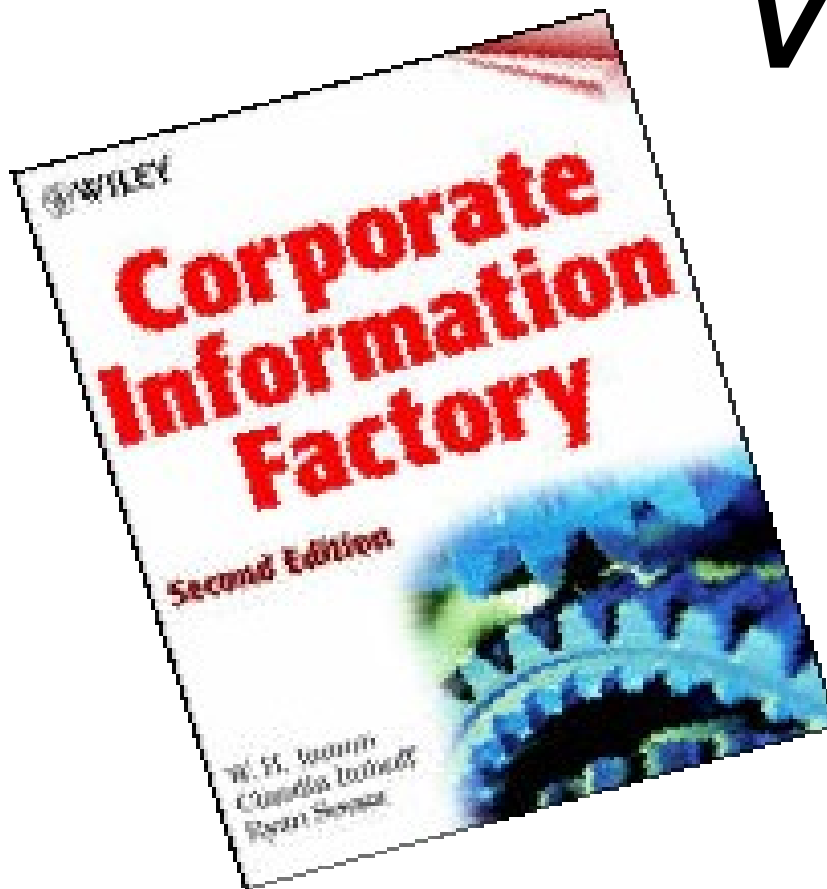
Data Consolidation

Captures data from multiple data sources and integrates into a single Master Data hub.

Data Federation

Provides a single virtual Master Data view of one or more data sources. No data is stored in the Master Data hub.

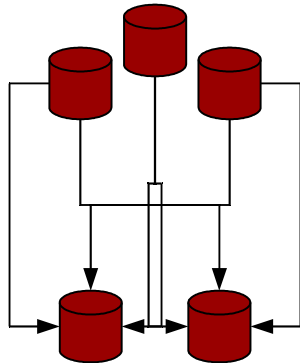
VS



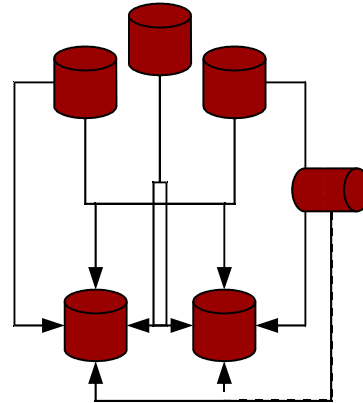
METAWRIGHT

Architecture choices

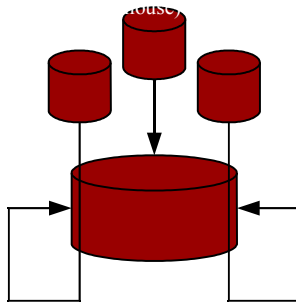
Independent Data Mart



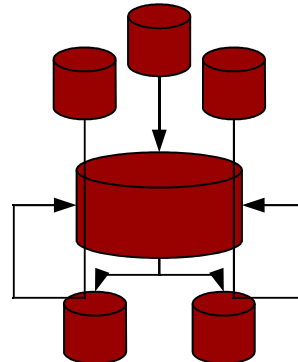
Data Mart Bus Architecture



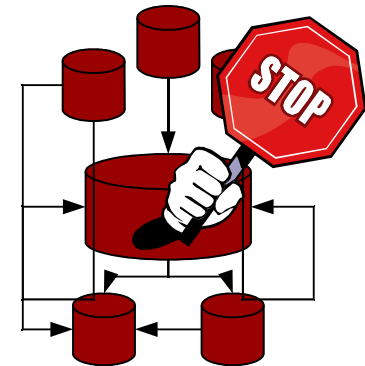
Centralized Data Warehouse



Hub & Spoke Architecture
(Corporate Information Factory)



Federated Data Stores



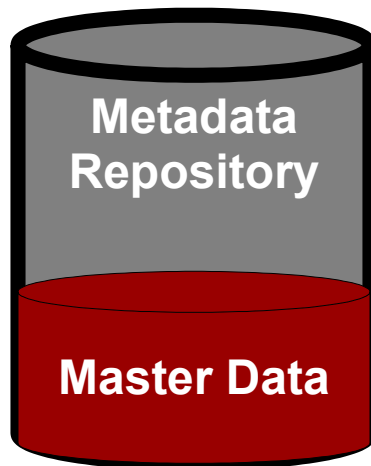
METAWRIGHT

Metadata + Repository

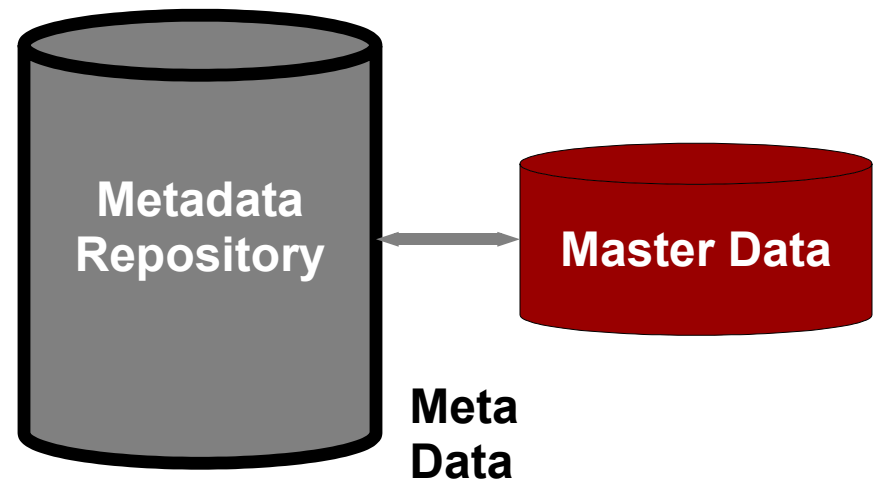
Are master data files metadata?

Need for central repository

**Inside Metadata
Repository**



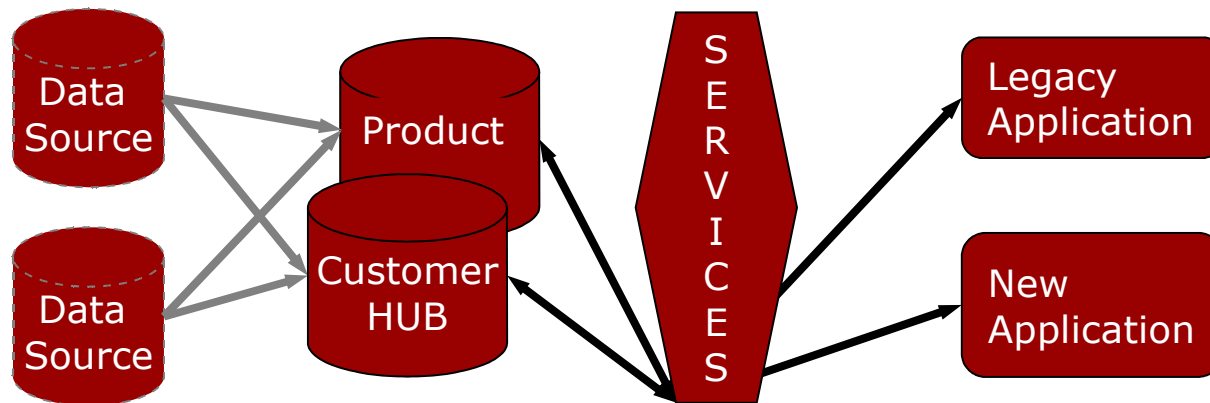
**Outside Metadata
Repository**



METAWRIGHT

Master Data Architecture – Repository

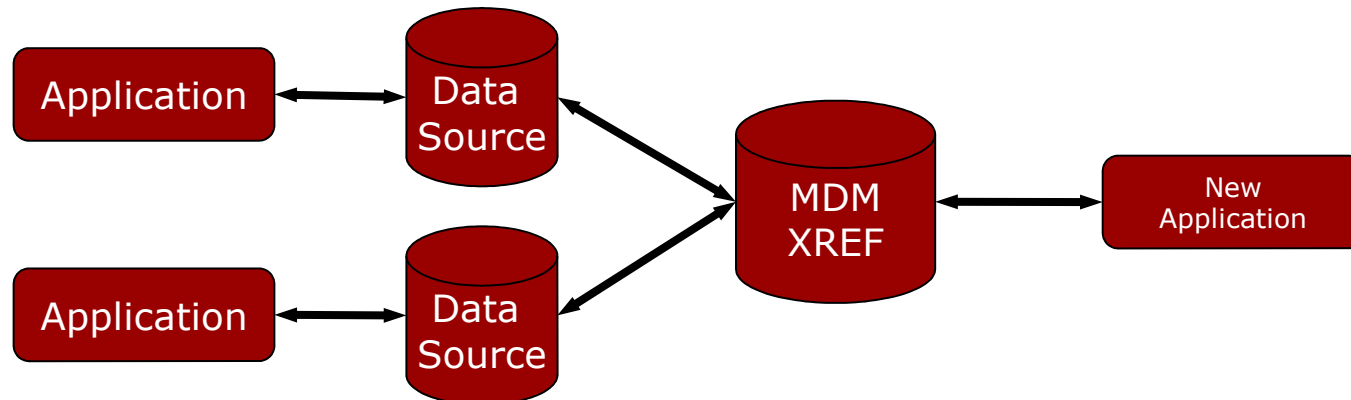
- **Comprehensive, single source for Master Data**
- **Centrally managed, highly effective data governance**
- **Fixed data model – typically Customer & Product**
- **Extensive and sophisticated data integration process**
- **All applications need to change to conform to Hubs**
- **Expensive to modify to new business requirements**
- **Hard to adapt to rapid changes**



METAWRIGHT

Master Data Architecture – Registry

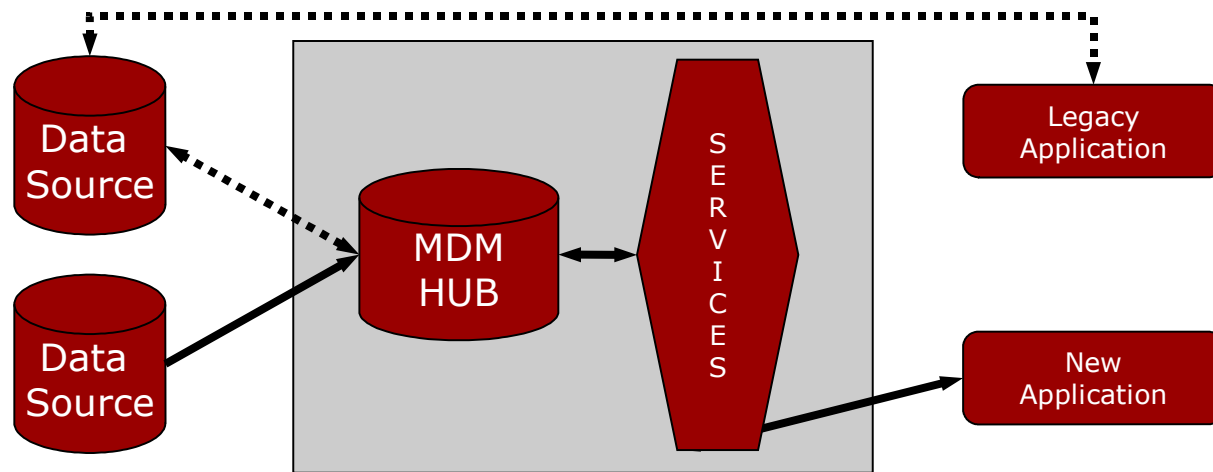
- **Linkages to other data stores with transform logic**
- **Low Cost, small footprint, only unique key kept in DB**
- **Minimal disruption of current applications**
- **Limited Data Governance, no merged Master Data**
- **No resolution of semantic disintegrity**
- **No History, System of Record linked to Data Source**
- **Query performance - data availability from multiple sources**



METAWRIGHT

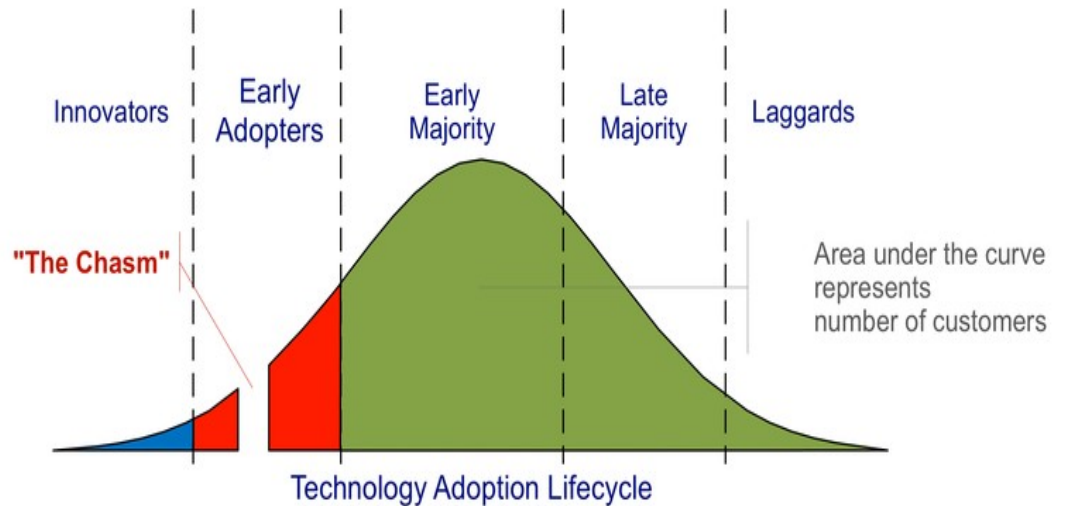
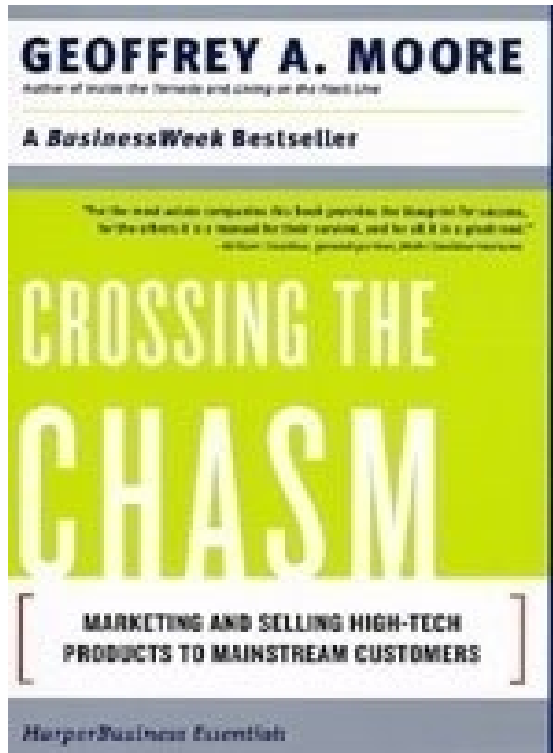
Master Data Architecture – Hybrid

- **Relatively low cost solution**
- **Extensible Data Model based on industry templates**
- **“Best we have” Master Data Model**
- **Generalized governance & maintenance facilities**
- **ETL to load HUB may be complex**
- **Data redundancy may lead to update conflicts and data latency issues**



METAWRIGHT

Technology Adoption



Product Life Cycle

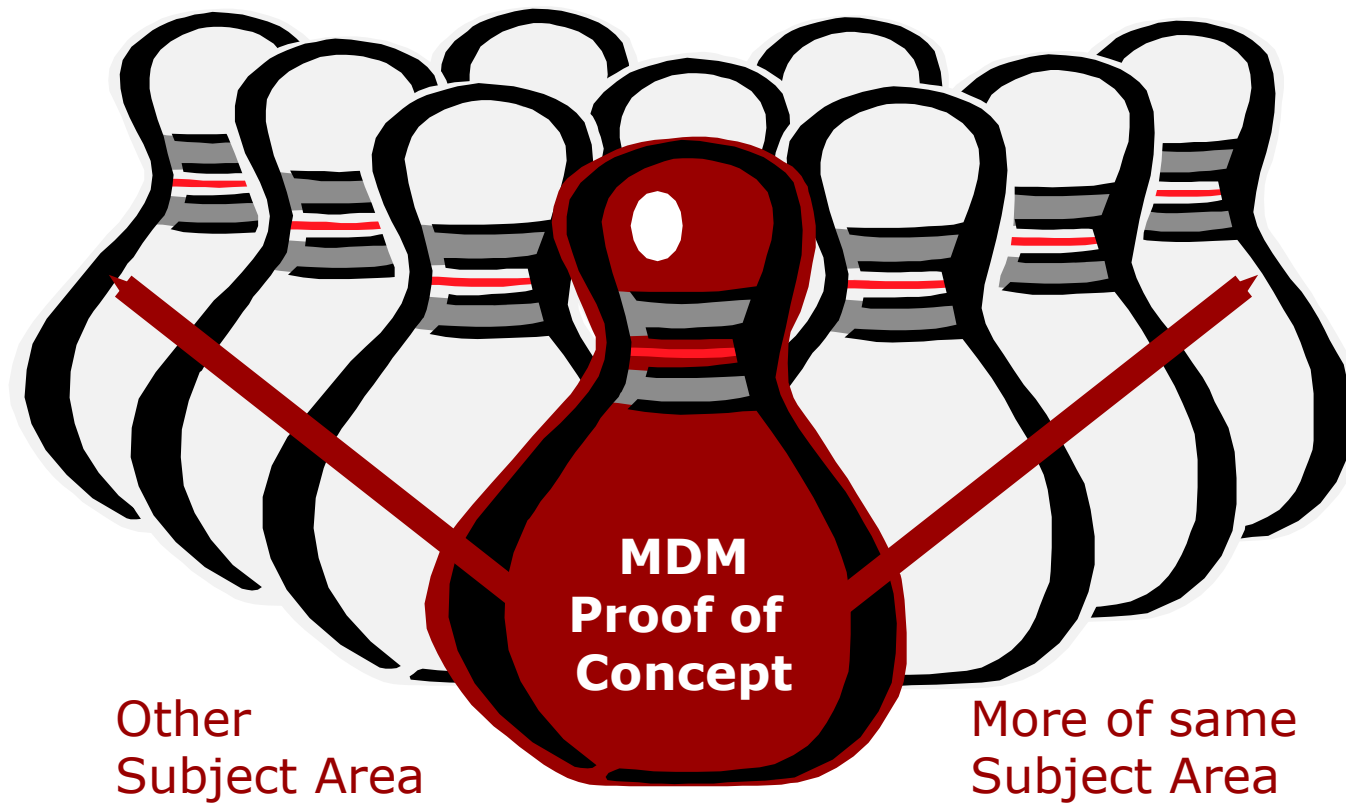
1. Market introduction

- **Costs are high**
- **Slow sales volumes to start**
- **Demand has to be created**
- **Customers have to be prompted to try the product**
- **Makes no money at this stage**

2. Growth

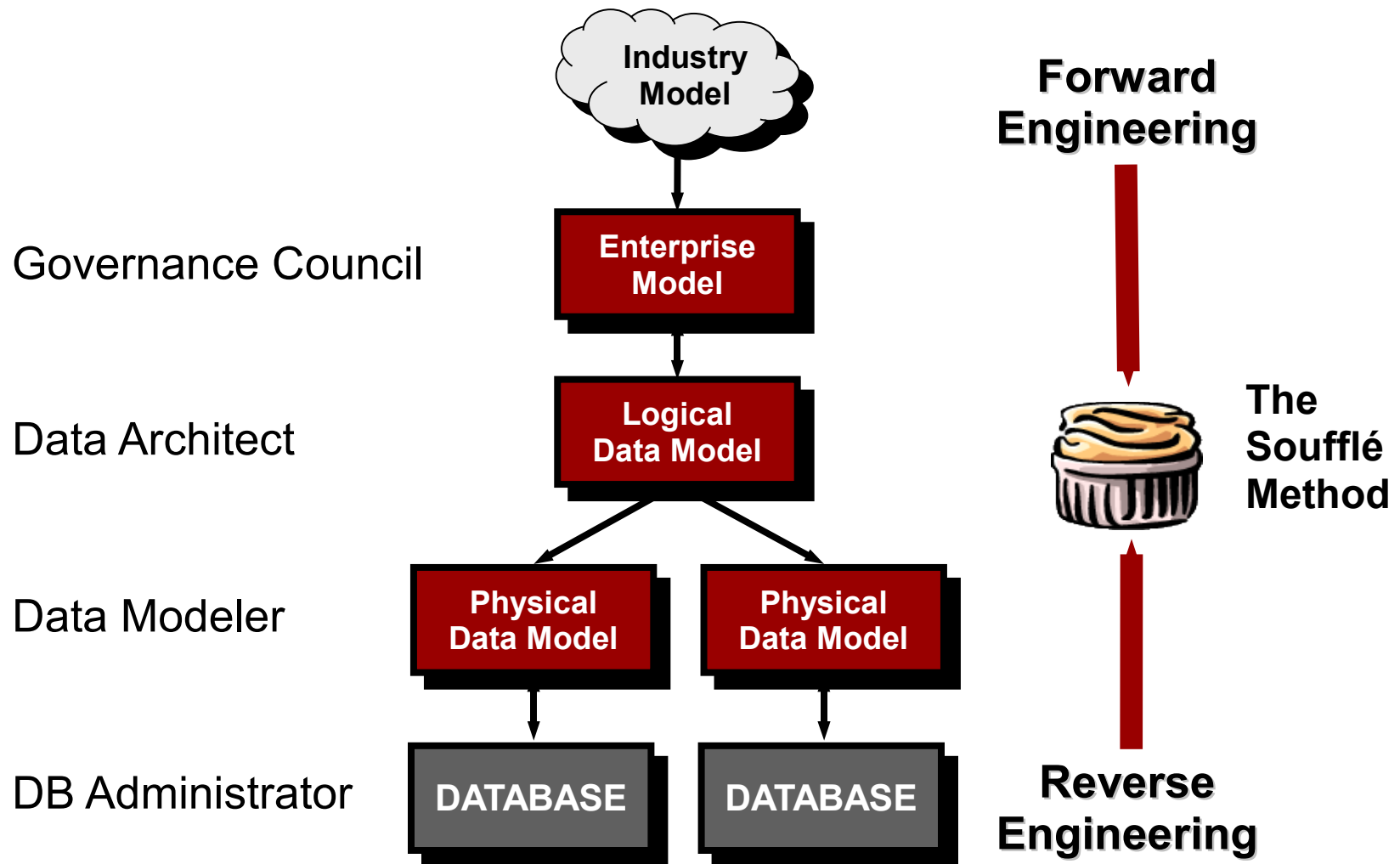
- **Costs reduced due to economies of scale**
- **Sales volume increases significantly**
- **Public awareness increases**
- **Profitability begins to rise**

Market Strategy



METAWRIGHT

Master Data Harmonization



METAWRIGHT

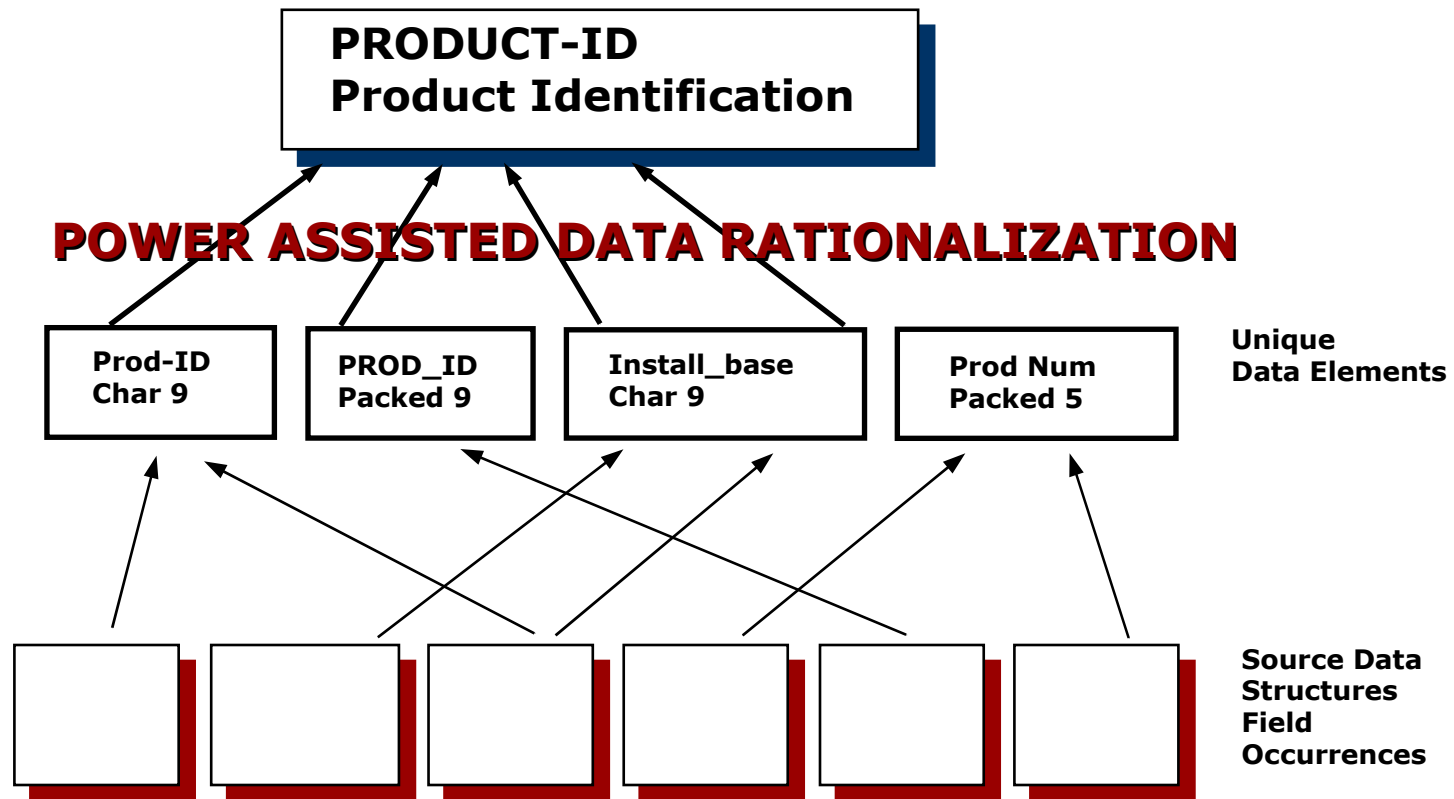
By any other name...



**Analytical
Master Data**

METAWRIGHT

Subject Area Integration



Data Modeling vs Data Profiling

Data Modeling documents metadata or columns

Customer Name	Company	Address	Year Purchased
Joe Smith	Komputers R Us	Germany	1970
Mary Jones	Big Bank Co	10 Gulf Road	1999
Proful Bishwal	Little Bank Inc	Manhattan	prospect
Ming Lee	My Favorite Store	PO Box 987	10/31/1984
Giovanni Petra	(*#&\$JKLJ	10 Main St	2000

Data Profiling analyzes instance data or rows

Traditional Meta-driven approach



Data Architects

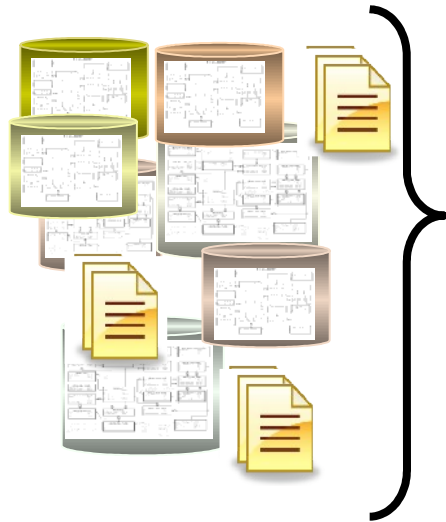
Design does not always match Reality!

Customer
Gen. Mot.
GM
9238423984
G-M 2342

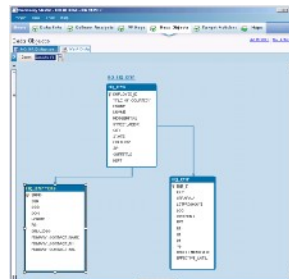
.



Data Discovery

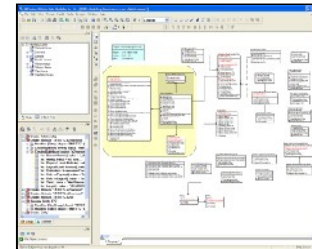


Data Profiler

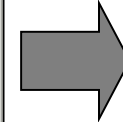


- Profile data
- Discover primary-foreign keys
- Identify orphan rows
- Find overlapping columns

Data Modeler



- Import inferred structures from Data Profiler
- Visualize information in an intuitive data model
- Integrate with other modeling efforts



Data Profiling – Match, Merge and Cleanse

Master Data Quality Analysis

Identify critical data elements

Review the cardinality, range, mode, and null rules

Review the value, pattern & length frequencies

Review parent-child relationships

Identify orphaned rows and values

Cross-System Analysis

Identify reference data overlap between data sources

Identify data content discrepancies between data sources

Validate mappings to MDM target

Master Data Error Types

Synonyms (False negatives) - two representations of the same entity are incorrectly defined as two separate entities

Appearing next month...

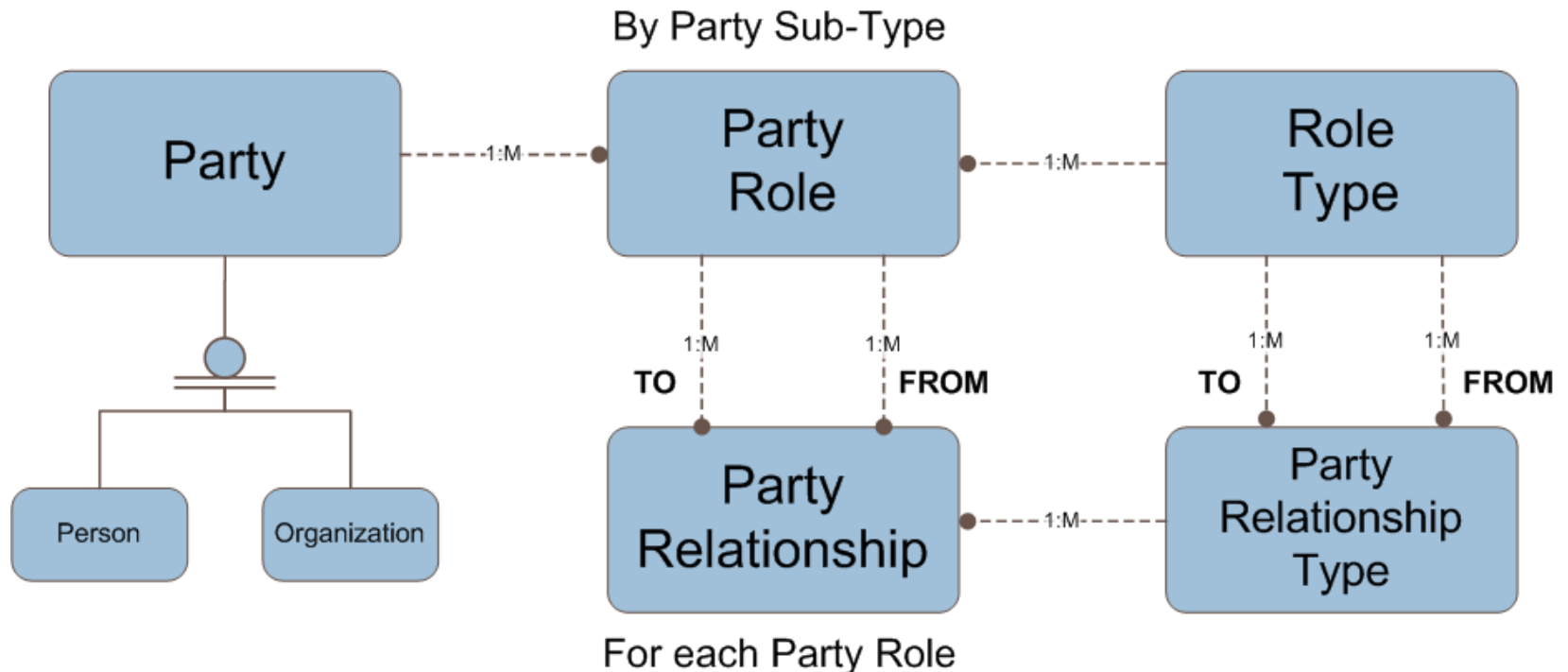
Why re-invent the wheel? that apply to well over *50 percent of most data model constructs* and that can be reused. It is this 50 percent that I address in this presentation.

Paul Agnew, Universal Patterns, how they can help you build your data models

November 19, 2009

Universal Models

Basic Party - Role



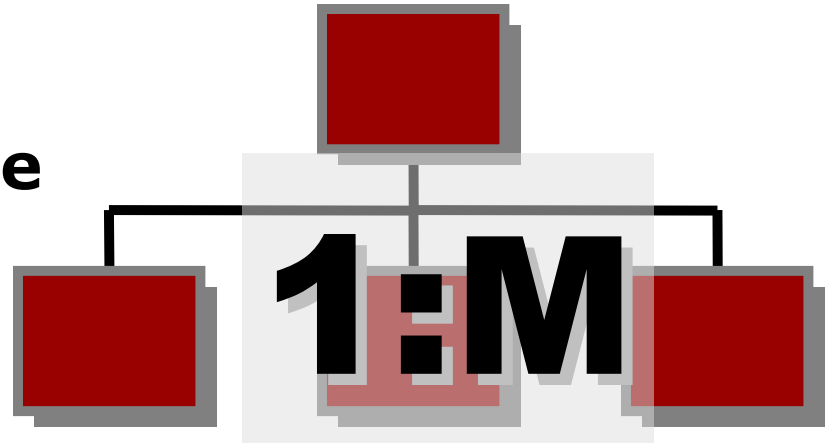
The Data Model Resource Book: Len Silverston

METAWRIGHT

Architecture Models

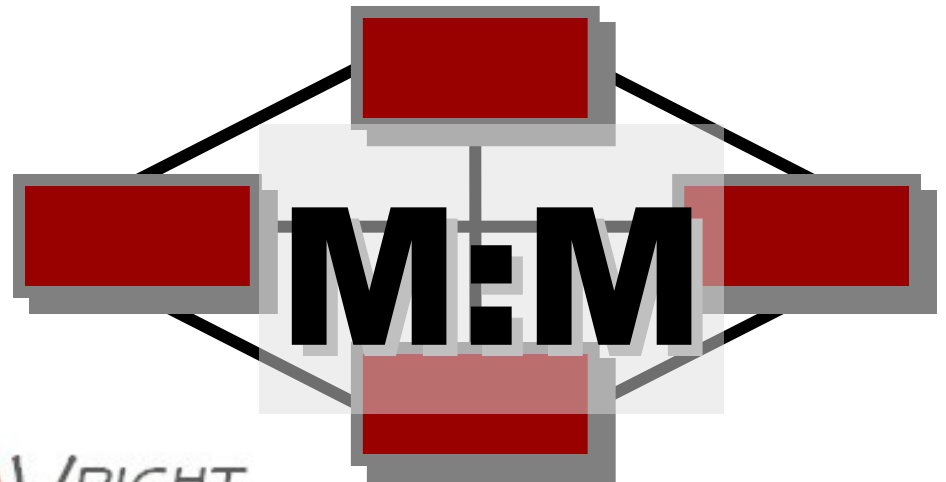
- **Process based**

Hierarchy mode



- **Topical or subject based**

Network model



METAWRIGHT

Data Governance

'The act or process of directing, leading and assuring that information is managed effectively as an enterprise resource, including resolving information conflicts, across the enterprise.'

Larry English

Data Governance – Building Trust in Data

People

- Data Stewards
- Data Architects
- Data Modelers

Process

- Data Security / Access
- Data Quality Assurance
- Master Data Management
- Metadata Management
- Enterprise Model Management

Data Quality Assurance

Data Stewardship

Data Profiling/Cleansing

System of Record – (Definition)

– Data Model

Versioning

Change Management

Health and Welfare

System of Reference – (Lifecycle)

– Workflow

Synchronization and Integration

Business Rules and Standards

External Referencing

METAWRIGHT



**You can't manage
what you don't
measure!**

METAWRIGHT

Scorecard – TDWI.org

Data Quality Dimensions	Completeness	Conformity	Accuracy	Consistency	Timeliness	Uniqueness	Redundancy	Duplication
Data Element List								
Obligor Identifier	100%	76%	94%	83%	95%	99%	97%	100%
Obligor Description	94%	85%	83%	96%	79%	99%	85%	86%
Agreement Identifier	80%	84%	84%	88%	87%	81%	79%	96%
Facility Default Grade	78%	76%	80%	77%	86%	95%	86%	87%
Obligor Default Grade	87%	90%	77%	97%	89%	87%	79%	94%
Probability of Default	100%	98%	93%	77%	75%	83%	96%	81%
Loss Given Default	87%	91%	83%	77%	86%	90%	97%	88%
Total Exposure Amount	79%	94%	90%	80%	76%	91%	89%	88%
Expected Loss Amount	97%	95%	98%	91%	94%	97%	79%	83%
Maturity Date	88%	82%	92%	81%	81%	88%	96%	96%
Collateralized Exposure Amount	88%	100%	84%	96%	79%	92%	89%	84%
Collateral Acquire Date	88%	86%	82%	87%	95%	95%	81%	84%

Data Stewardship

- **Difficult to assign,
not process based**
- **Cross-functional**
- **Not a Single DATA
Steward**

**Business,
Technical,
Operational
Accountability,
Authority,
Responsibility**

Structure vs. Content

		Columns		
Rows		Cells		

Master Data Stewardship

Difficult to assign, not process based

Infrastructure — cross-functional

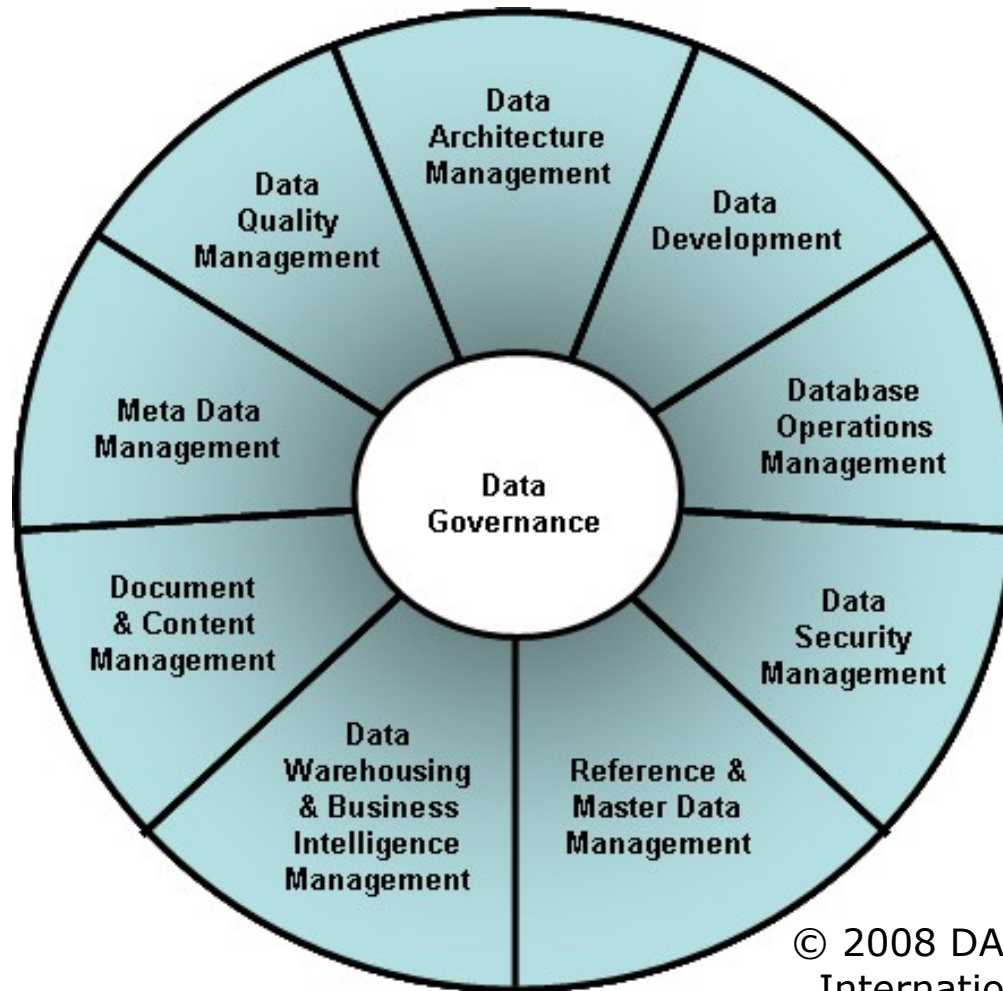
Business stewards - Content

Descriptions

Technical stewards - Structure

Codes and structure

Data Management Body of Knowledge



© 2008 DAMA
International

METAWRIGHT

References

**Managing Reference Data in Enterprise Databases:
Binding Corporate Data to the Wider World**

Malcolm Chisholm - Morgan Kaufmann

**The Data Model Resource Book: A Library of Universal
Data Models for All Enterprises (VOL. 1,2)**

Len Silverston - John Wiley & Sons

**The Data Model Resource Book: Universal Patterns for
Data Modeling**

Len Silverston - John Wiley & Sons

**Master Data Management & Customer Data Integration
for a Global Enterprise**

Alex Berson & Larry Dubov - McGraw Hill



Thank You!

Donald J. Soulsby
DSoulsby@MetaWright.com



META^WRIGHT