



Whitemarsh
Information Systems Corporation

WRAD Conference 2000 Cast a Wide Net

Achieving Enterprise Wide
Data Semantics Standardization
in Support of Data Warehousing

Whitemarsh Information Systems Corporation
2008 Althea Lane
Bowie, Maryland 20716
Tele: 301-249-1142
Email: mmgorman@wiscorp.com
Web: www.wiscorp.com

Table of Contents

1.	Dreadful Information Technology Statistics (Standish 1995 & 1999)	1
2.	Data Is Executed Policy	3
3.	“Database Objects” Exist Within a Knowledge Worker Framework	8
4.	Data (Policy & Procedures) Exists Across 5 Data Architecture Classes	11
6.	The Data Standardization Work Plan	22
7.	Financial Benefits from Standardizing Data	28
8.	Summary	29



1. Dreadful Information Technology Statistics (Standish 1995 & 1999)

Result Classification	Percent	
	Utility (Gas and Electricity) Companies	365 Respondents on the Success of Client/Server Development Efforts
Succeed	24	16
Challenged	43	53
Failed	33	31

Legend	Category Description
Succeed	On-time, within budget, features as promised
Challenged	Late, greater than budgeted, less features than promised
Failed	Cancelled outright



Top Three Reasons for Success

- ! User Involvement
- ! Executive Management Support
- ! Clear Statement of Requirements

Top Three Reasons for Failure

- ! Incomplete Requirements
- ! Lack of User Involvement
- ! Lack of Resources



2. Data Is Executed Policy

- ! Data is “what” remains after *Policy* is executed.
- ! Processes are the *Procedures* through which the policy execution is accomplished
- ! “Data” Definitions are thus Policy Definitions
- ! Data Administration, Data Standardization, and Database Design is a “Enterprise Policy Specification” activity

To have *Complete* database is to have *BOTH* defined policy and supporting procedures...

- ! Each table is the data representation of a policy's definition
- ! The rules by which rows are added, deleted, and modified are the mechanisms necessary for policy execution, that is, the policies procedures
- ! A fully defined policy includes both its complete definition and its necessary steps for coherent execution
- ! Interrelated collections of rows form more comprehensive policies

We must be about Complete Database, that is, Database Objects



DBMS and the History of Policies and Procedures

- ! Complete database DBMS from 1969 thru the mid 1980s via CODASYL DBMS which had both integrated Policy & Procedures
- ! But relational DBMS (SQL:1986-SQL:1992) exalted data (Policy) but trashed processes (Procedures)
- ! Now finally, SQL:1999 has re-joined Policy (data) & Procedures (process) into a single ANSI DBMS standard¹

¹ More on SQL:1999 can be obtained from www.wiscorp.com in the article and presentation, “Great News, The Relational Model is Dead.”



When Only Data is Defined

Key Questions regarding data distribution effects	Semantic Control			
	Centralized		Decentralized	
	Data Storage Control			
	Centralized	Decentralized	Centralized	Decentralized
Is data able to be shared among sites?	yes	yes	no	no
Is concurrent processing of the same data possible?	yes	maybe	no	no
Are common or corporate reports possible?	yes	yes	no	no
Can there be an overbearing "big brother" feeling?	yes	maybe	no	no
Is there local control and ownership?	no	maybe	yes	yes
Does there need to be common data standards & policies?	yes	yes	no	no
Can there be common and enforced business rules?	no	no	no	no
Can local data requirements be satisfied?	maybe	yes	maybe	yes



When Only Processes are Defined

Questions regarding program/system distribution effects	Development Control			
	Centralized		Decentralized	
	Execution Location			
	Centralized	Decentralized	Centralized	Decentralized
Is the same program able to be shared among sites?	yes	yes	no	no
Is concurrent processing of the same data possible?	yes	maybe	no	no
Are common or corporate reports guaranteed?	yes	maybe	no	no
Can there be an overbearing "big brother" feeling?	yes	maybe	no	no
Is there local control and ownership?	no	maybe	yes	yes
Does there need to be common processing standards & practices?	yes	yes	no	no
Can there be common and enforced data standards?	no	no	no	no
Can local processing requirements be satisfied?	maybe	yes	maybe	yes



But when Both Policy (data) and procedures (processes) are defined through Database Objects.

Questions regarding database object distribution effects	Semantic Control			
	Centralized		Decentralized	
	Development Control			
	Centralized	Decentralized	Centralized	Decentralized
Are database objects able to be shared among sites?	yes	yes	no	no
Is concurrent processing of the same database object instance possible?	yes	maybe	no	no
Are common or corporate reports possible?	yes	yes	no	no
Can there be an overbearing "big brother" feeling?	yes	maybe	no	no
Is there local control and ownership?	no	maybe	yes	yes
Does there need to be common data standards & policies for database objects?	yes	yes	no	no
Can there be common and enforced business rules as well as standard data semantics?	yes	yes	yes	no
Can local data requirements be satisfied?	maybe	yes	maybe	yes



3. “Database Objects” Exist Within a Knowledge Worker Framework

Knowledge Worker Framework						
Deliverables	Mission	Man-Machine Interface				
		Machine		Interface	Man	
		Database Object	Business Information System	Business Event	Business Function	Organization
Scope	List of business missions	List of major business resources	List of business information Systems	List of interface events	List of major business scenarios	List of organizations
Business	Mission hierarchies	Resource life cycles	Information sequencing and hierarchies	Event sequencing and hierarchies	Business scenario sequencing and hierarchies	Organization charts, jobs and descriptions
System	Policy hierarchies	Specified data model and Database object models	Information system designs	Invocation protocols, input and output data, and messages	Best practices, quality measures and accomplishment assessments	Job roles, responsibilities, and activity schedules
Technology	Policy execution enforcement	Implemented Data Model	Information systems application designs	Presentation layer information system instigators	Activity sequences to accomplish business scenarios	Procedure manuals, task lists, quality measures and assessments
Deployment	Installed business policy and procedures	Operational Data Model	Implemented information systems	Client & server windows and/or batch execution mechanisms	Office policies and procedures to accomplish activities	Daily schedules, shift and personnel assignments
Operations	Operating business	Application Interface Data Models	Operating information systems	Start, stop, and messages	Detailed procedure based instructions	Daily activity executions, and assessments



Why is the Knowledge Worker Framework Important ?

! United States Government Accounting Office's Allocation of Information Technology Failures

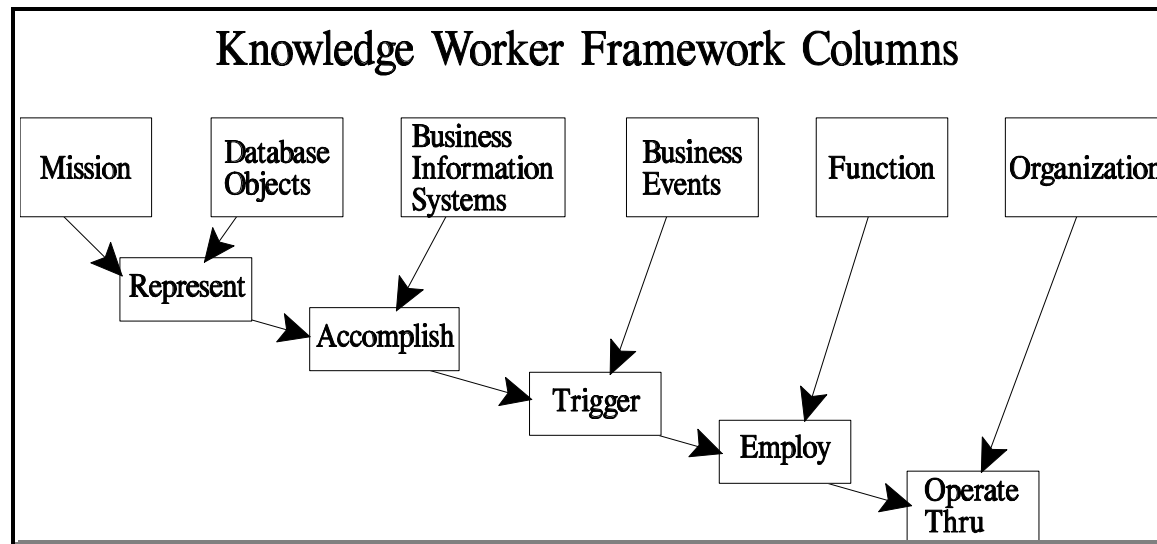
Knowledge Worker Framework							Row Totals of GAO Allocated Errors in Percent
Deliverables	Mission	Man-Machine Interface					
		Machine		Inter-face	Man		
		Database Object	Business Information System	Business Event	Business Function	Organ- ization	
Scope	5	2	3	1	3	4	18
Business	5	3	2	1	6	6	23
System	3	2	2	1	12	8	28
Technology	1	0	0	0	8	6	15
Deployment	0	0	0	0	5	5	10
Operations	0	0	0	0	3	3	6
Col. Totals	14	7	7	3	37	32	100

Note: All numbers expressed as Percent allocations of errors to cells

12 Gray cells are IT Cells

If IT were Zero Percent Failure the 95% of all IT Systems would still fail

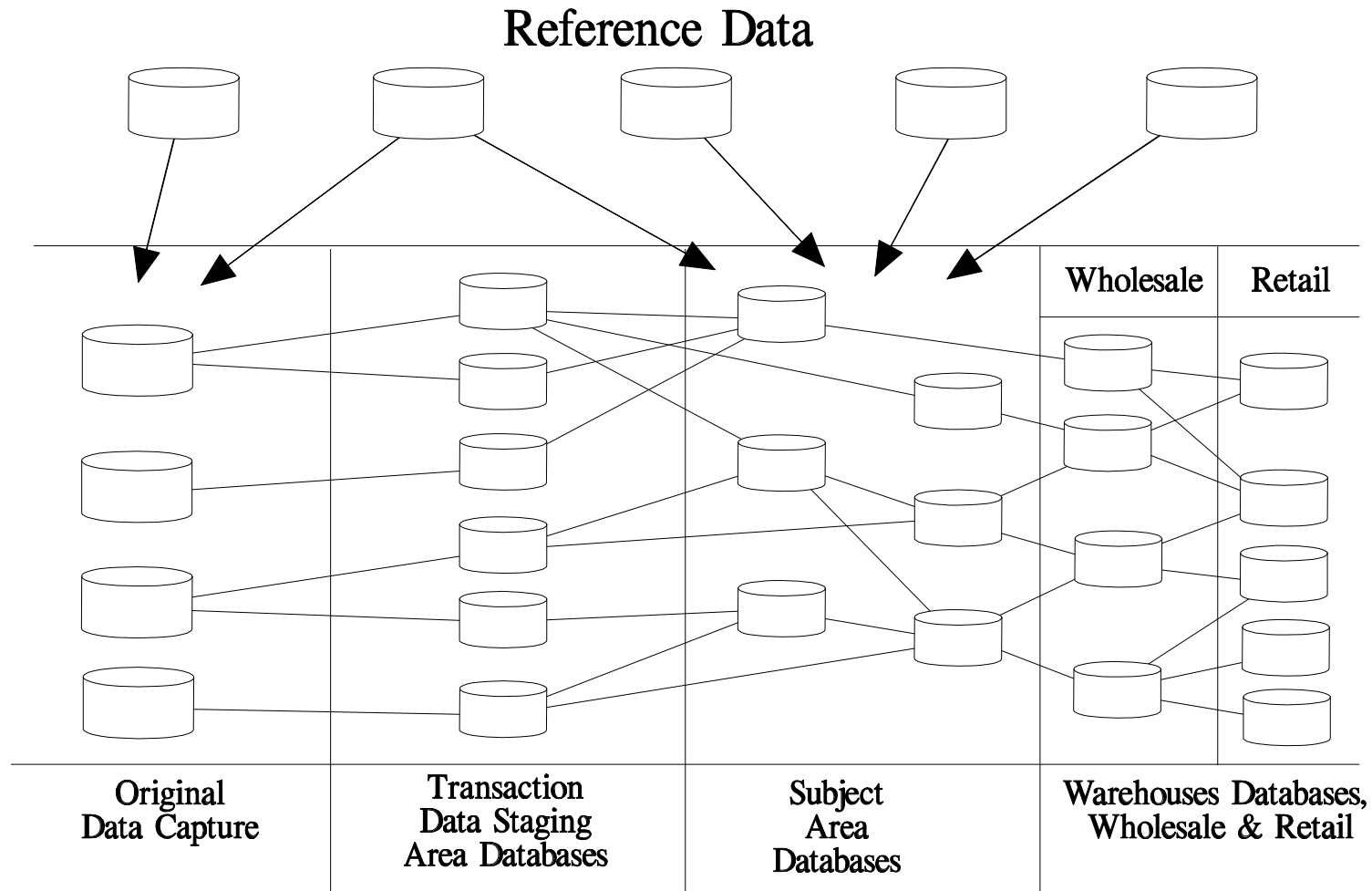




Knowledge Worker Framework						
	Mission	Database Object	Business Information System	Business Event	Function	Organization
Scope	To be successful at Knowledge Worker Management, all rows and columns must be integrated through:					
Business						
System						
Technology						
Deployment						
Operations						



4. Data (Policy & Procedures) Exists Across 5 Data Architecture Classes



Data Architecture Class		Key Characteristics	
Reference Data		Durable codes and long value alternatives Company-wide standard semantics Source of all valid and invalid values Changed data history supported by conversion mappings	
Original Data Capture		Original business data Captured at source Application specific Vendor package if possible	
Transaction Data Staging Area		Business transaction data Transformed to common format Application specific Custom, but simple applications	
Subject Area Databases		Integrated business data Broad and comprehensive Subject area coverage Custom, but simple	
Data Warehouse	Wholesale	End-user business data Specific needs design Application specific	<i>Data Warehouse Quality is INFINITELY proportional to the Multiplicative Sum of all previous data qualities. If all previous is GREAT, DWH is OK. If all previous is bad then DWH is UNBELIEVABLY BAD!</i>
	Retail	Vendor package if possible	



5. Achieving Data Standardization

Goal: Where ever practical, anything that is the same should:

- ! Mean the same however it is named or deployed
- ! Have the same data semantic characteristics

These Business Facts are the same !

Estimated North East Sales Amount USD	5
---------------------------------------	---

Accuracy	Estimated
Geography	North East
Sales Amount	5
Currency	USD

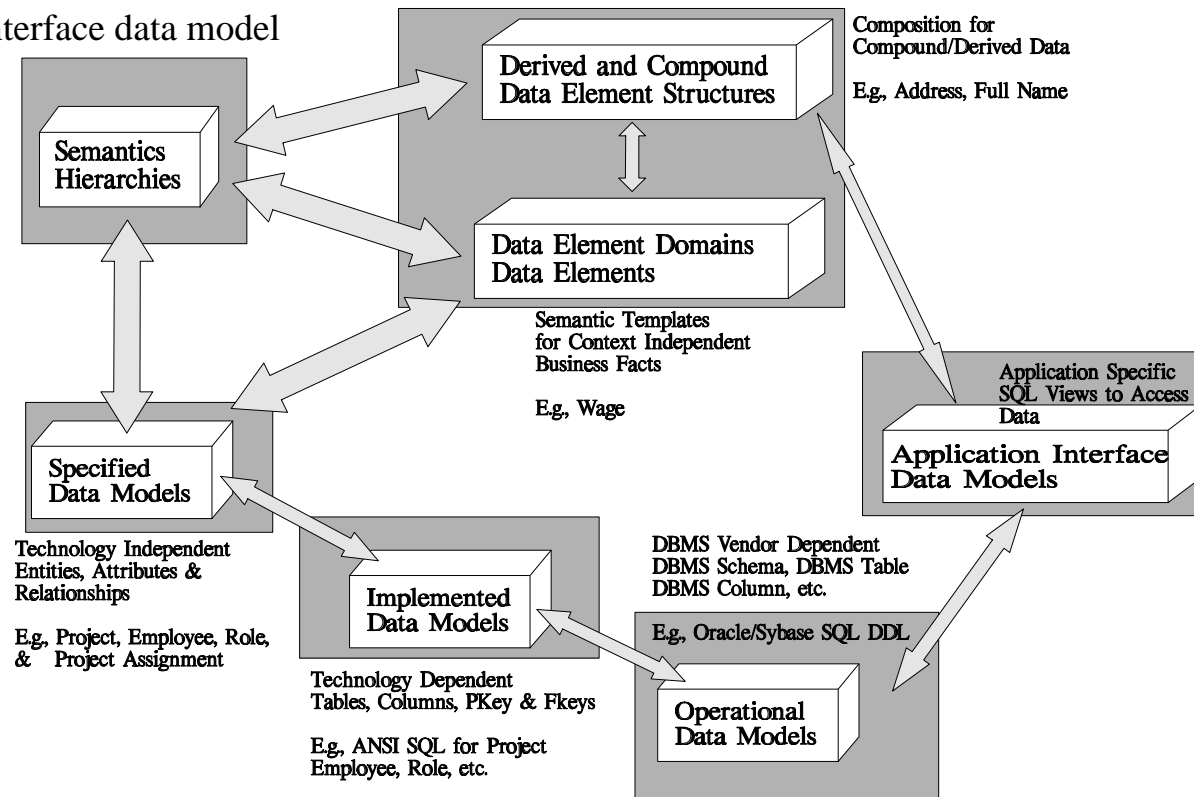
Regardless of whether the business fact is resident in:

- ! Original data capture
- ! Transaction data staging area
- ! Subject area database
- ! Data warehouse (wholesale or retail)
- ! Reference data

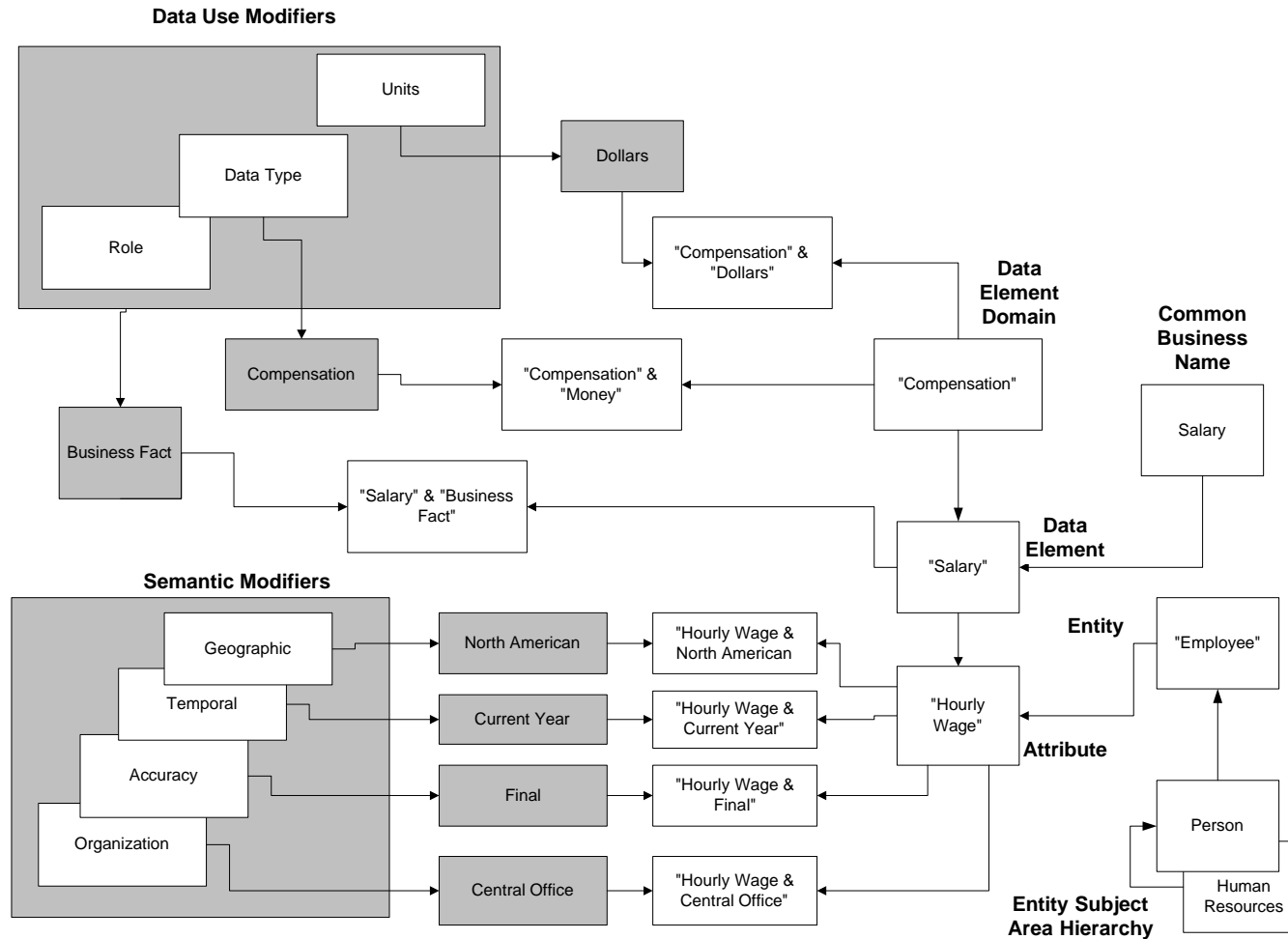


Key Meta-models Involved in Data Standardization

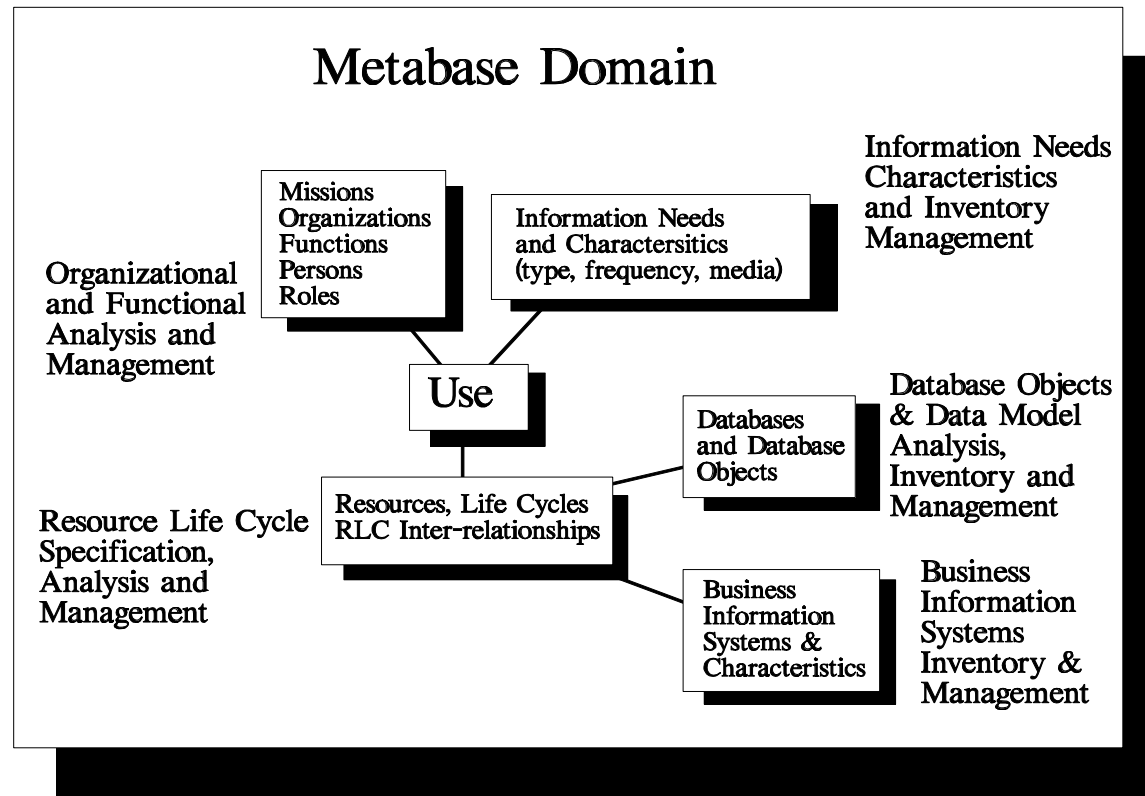
- ! Semantic hierarchies
- ! Data elements
- ! Specified data models
- ! Implemented data models
- ! Operational data models
- ! Application interface data model



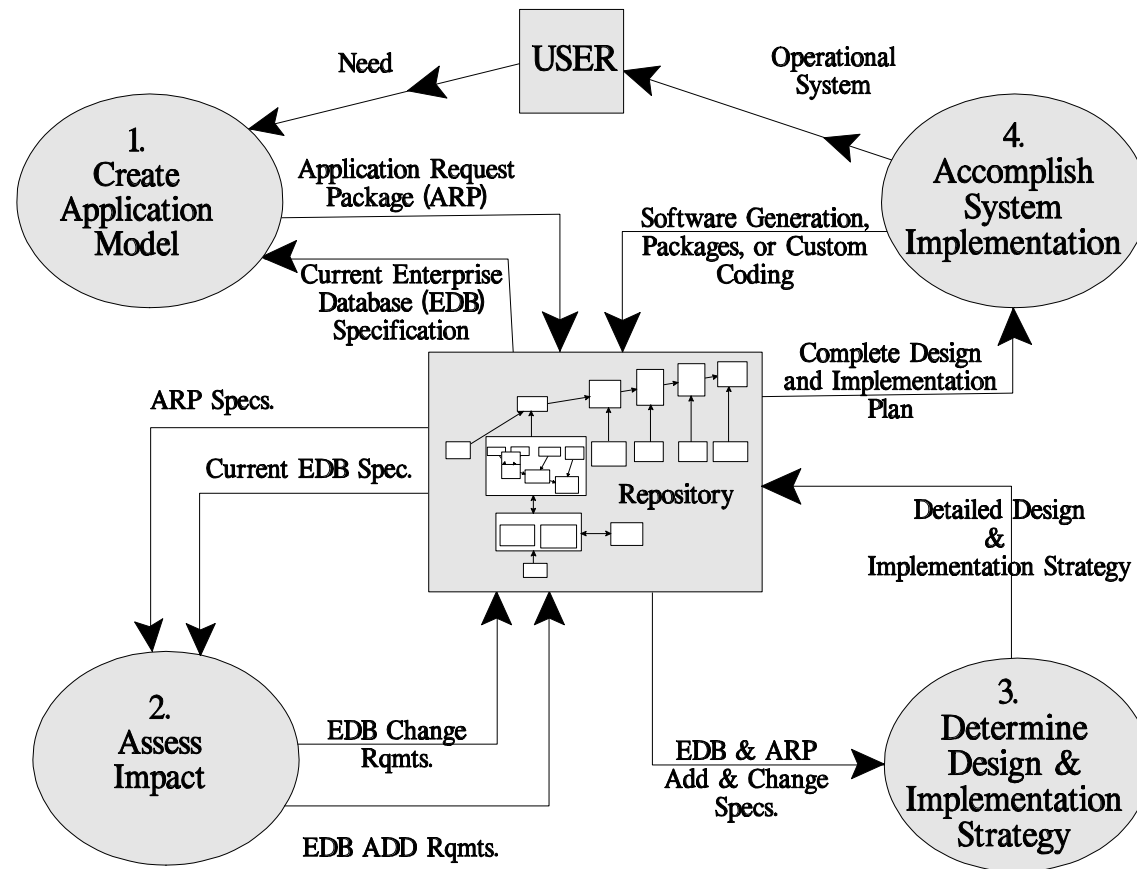
Employing Semantics to Data Element Domain, Data Element, and Attributes



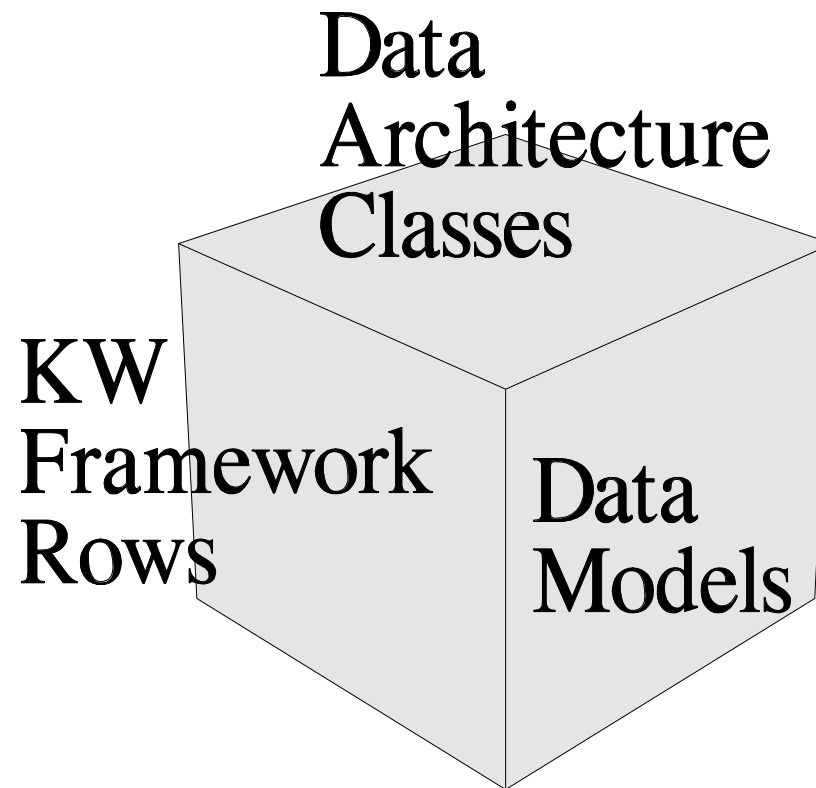
Data Standardization Metadata Data Models Exist Within an Overall Metabase



The Metabase Exists Within an Overall Information Technology Environment



There are Key Intersections Among...



Data Architecture Classes and Knowledge Worker Framework Rows

Knowledge Worker Framework Row	Data Architecture Classes					
	Reference Data	Original Data Capture	Transaction Data Staging Area	Subject Area Databases	Data Warehouses	
					Wholesale	Retail or Data Marts
Scope	<i>Every database and database project from every data architecture class must go through every Knowledge Worker Framework row</i>					
Business						
System						
Technology						
Deployment						
Operations						

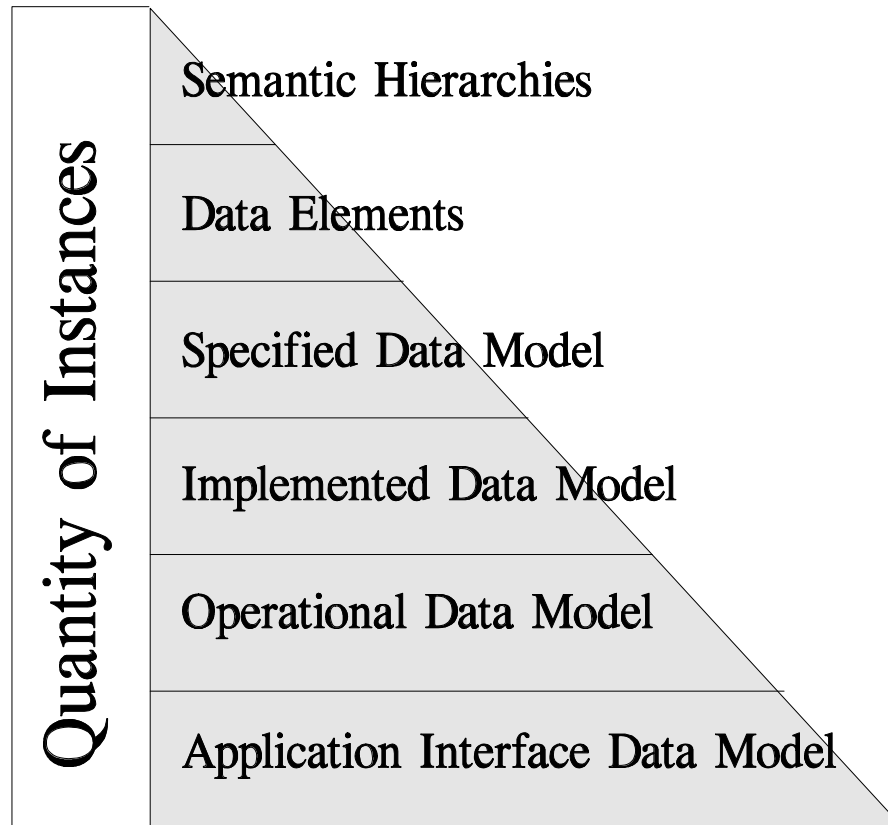


Data Architecture Classes and Instances of Data Models

Instances of Data Models	Data Architecture Classes					
	Reference Data	Original Data Capture	Transaction Data Staging Area	Subject Area Databases	Data Warehouses	
					Wholesale	Retail or Data Marts
Semantics	Absolutely and unconditionally, only one across the entire enterprise					
Data Elements	As few unique data elements within each database architecture class as possible					
Specified	As few as is possible for each different policy triple: { subject, entity, and attribute }					
Implemented	One for each database	One for each database	One for each database	One for each database	One for each database	One for each database
Operational	One for each deployed DBMS database	One for each deployed DBMS database	One for each deployed DBMS database	One for each deployed DBMS database	One for each deployed DBMS database	One for each deployed DBMS database
Application Interface	A view for each database & program interface	A view for each database & program interface	A view for each database & program interface	A view for each database & program interface	A view for each database & program interface	A view for each database & program interface



Uniquely Database: Define Once, Use Many Times



6. The Data Standardization Work Plan

Phase 1. Data Standardization Project Planning, which provides the work breakdown structure for planning data standardization projects

- ! Perform overall data standardization project planning
 - , Determine and achieve consensus on overall goals and objectives
 - , Determine and achieve consensus on success and evaluation measures
 - , Review, revise, and achieve consensus on work breakdown structure (WBS)
 - , Identify business unit involvement in project and phases
 - , Identify and assign staff and accomplish project estimates
 - , Acquire computing environment and training

- ! Review, revise, and achieve consensus on all deliverable content and format for all comparison, difference, and resolution reports including
 - , Identified “data element”²s
 - , Business domains of “data element”s
 - , Existing “data element” characteristics such as business domain, and other semantics
 - , Enterprise standard (international, regional, national, and local) “data element” characteristics
 - , Difference between existing and enterprise standard

- ! Develop resolution mechanism for “data element” semantic differences including
 - , Inter business unit automation interaction
 - , Inter business unit human communication interaction
 - , Estimated resources (hardware, software, peopleware, and time) required to resolve semantic differences
 - , Identify and quantify business risk and/or impact associated with unresolved differences

² Here, *data element* refers to data element domains, data elements, attributes, columns, DBMS columns, or view columns as applicable during the data standardization projects. A data element is a semantic template, not a value bearing fact.



Phase 2 Data Standardization “data element” Identification and Analysis, which provides the work breakdown structure for identifying the critical “data element”s that are candidates for data standardization

- !** Identify, assign staff and accomplish phase estimate
 - , Identify phase project manager and determine members of phase team
 - , Develop detailed phase estimate
 - , Accomplish resource loading, build PERT, Gantt and CPM charts for phase
 - , Present phase plans and revise as necessary
 - , Identify and assign administrative support
 - , Identify and acquire automation/tools support

- !** Identify or develop mission and perform analysis
 - , Identify or create overall mission for data standardization area
 - , Create appropriate subordinate missions relevant to the data standardization area
 - , Create mission for data standardization area
 - , Create, store and validate through reporting the mapping between enterprise standard missions and the missions of the data standardization area
 - , Create mission comparison, differences, and resolution report
 - , Analyze report, rank issues, and make assignments for differences resolution
 - , Identify relevant database domains

- !** Identify or develop database domains and perform analysis
 - , Select or create appropriate database domains
 - , Create appropriate subordinate database domains relevant to the area of the data standardization area
 - , Create database domains for data standardization area
 - , Create, store and validate through reporting the mapping between enterprise standard database domains and the database domains of the data standardization area
 - , Create database domain comparison, differences, and resolution report
 - , Analyze report, rank issues, and make assignments for differences resolution
 - , Identify relevant database objects



- ! Identify or develop database objects and perform analysis
 - , Select or create appropriate database objects
 - , Analyze the scope, purpose and coverage of the data standardization area's database objects
 - , Create, store and validate through reporting the mapping between enterprise standard database objects and the database objects of the data standardization area
 - , Create database object comparison, differences, and resolution report
 - , Analyze report, rank issues, and make assignments for differences resolution

- ! Identify or develop database object data structure analysis
 - , Select or create appropriate database object data structures
 - , Analyze the scope, purpose and coverage of the data standardization area's database object data structures as evidenced through its tables and/or files
 - , Create, store and validate through reporting the mapping between enterprise standard database object data structures and the database object data structures of the data standardization area
 - , Create database object data structure comparison, differences and resolution report
 - , Analyze report, rank issues, and make assignments for differences resolution

- ! Perform standard "data element" analysis
 - ! Select or create appropriate "data element"s including
 - ! Select or create semantics for "data element"s
 - ! Identify data standardization area's "data element" deployment
 - ! Create, store and validate through reporting the mapping between standard "data element"s and the deployed uses of the "data element"s within the data standardization area
 - ! Create a comparison, difference, and resolution report for international, regional, national, and local "data element"s
 - ! Analyze report, rank issues, and make assignments for differences resolution



- ! Perform deployed “data element” analysis
 - , Select or create appropriate deployed “data element”s including
 - , Select or create semantics for deployed “data element”s for
 - , Identify data standardization area’s deployed “data element”s
 - , Create, store and validate through reporting the mapping between standard “data element” and the deployed “data element”s within the data standardization area
 - , Create a comparison, difference, and resolution report for international, regional, national, and local deployed “data element”s including:
 - , Analyze report, rank issues, and make assignments for differences resolution

- ! Perform business policy research and formulation
 - , Identify current policy basis for the deployed “data element”

- ! Review and/or formulate data definitions and standards for deployed “data element”s
 - , Identify all deployed “data element” that are intended to embrace the same set of semantics and values by business unit, computing system, and database
 - , Analyze the semantics within the deployed “data element” set to determine differences
 - , Analyze the values within sets of deployed “data element”s to determine differences
 - , Determine the mapping between semantic differences for deployed “data element”s differences resolution
 - , Determine the mapping between value set differences deployed “data element”s



Phase 3 Data Standardization Implementation, which provides the work breakdown structure for determining the what needs to be done to achieve data standardization for a particular “data element” and for its deployment throughout its implemented set of data fields.

- ! Identify, assign staff and accomplish phase estimate
 - , Identify phase project manager
 - , Determine members of phase team
 - , Develop detailed phase estimate
 - , Accomplish resource loading, build PERT, Gantt and CPM charts for phase
 - , Present phase plans and revise as necessary
 - , Identify and assign administrative support
 - , Identify and acquire automation/tools support

- ! Accomplish “data element” Standardization
 - , Obtain the report from “data element” standardization
 - , Determine the resources (hardware, software, peopleware, and time) for required policy changes
 - , Determine the resources required for operating system changes
 - , Determine the resources required for existing database changes
 - , Determine the resources required for historical data system changes
 - , Determine the intra and inter business unit risk for not accomplishing data standardization

- ! Assignment of management for review and approval
 - ! Present the report that identifies, analyzes, and costs the effect of nonstandardization of critical “data element”s
 - ! Present the report that identifies the costs of a critical “data element” data standardization effort
 - ! Present the recommendation for data standardization
 - ! Obtain a management decision to remain the same or to proceed with data standardization effort



- ! Proceed with data standardization implementation project
 - , Plan project
 - , Accomplish the database changes
 - , Accomplish the existing computing system changes
 - , Accomplish the policy and procedure changes
 - , Accomplish the actual data value migration for current and historical data
 - , Deliver the project results and develop lessons learned



7. Financial Benefits from Standardizing Data

Activity	Quantity	Cost via technique employed for definition
Starting quantity of columns/fields	19,000	\$6.75 million
Elimination of closely named columns and fields reduced the quantity to	3,000	\$1.06 million
Elimination of same concept but very differently named columns and fields reduced the quantity to	560	\$200,000



8. Summary

Topic	Summary Point
What is Data	Coherent unity of Policy and Procedures
Source of Information Technology Errors	<i>GAO States: ONLY 5% IT..... 95% Non-IT</i>
What is “Complete” Database	Specified through operational data models with BOTH data and process (a.k.a, policy & procedures)... In short, database objects completely within the firewall of the SQL:1999 compliant DBMS
Data Architecture Classes	Reference Data Original data capture Transaction Data Staging Area Subject Area Database Data warehouse, both wholesale and retail (data marts)
Data Warehouse Quality	<i>Data Warehouse Quality is INFINITELY proportional to the Multiplicative Sum of all previous data qualities. If all previous is GREAT, Data Warehouse is OK. If all previous is bad then Data Warehouse is UNBELIEVABLY BAD!</i>



Topic	Summary Point	
Key Metadata Models Essential for Data Standardization	Semantics	Absolutely and unconditionally, only one across the entire enterprise
	Data Elements	As few unique data elements within each database architecture class as possible
	Specified	As few as is possible for each different policy triple: {subject, entity, and attribute}
	Implemented	One for each database
	Operational	One for each deployed DBMS database
	Application Interface	A view for each database & program interface
Data Standardization Work Plan	Project Planning, which provides the work breakdown structure for planning data standardization projects	
	“Data Element” Identification and Analysis, which provides the work breakdown structure for identifying the critical “data elements” that are candidates for data standardization	
	Implementation, which provides the work breakdown structure for determining the what needs to be done to achieve data standardization for a particular “data element” and for its deployment throughout its implemented domain.	

