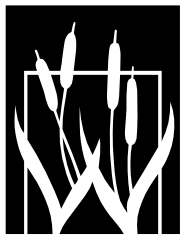


Whitemarsh
Information Systems Corporation

THE Problem
in
Enterprise Wide Data Standardization

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

Topics

1.0 Frameworks 1

2.0 Data Standardization, The Key To Interoperability 5
2.1 Typical Data Standardization "Problem" 8
2.2 Data Standardization is not just an abstract concept... 10
2.4 Common Reasons for Data Standardization Failures 12
2.4.1 Failure Reason: Having a Fundamentally Flawed Paradigm 13
2.4.2 Failure Reason: Not Dealing with Enterprise-Wide Data Architectures 16
2.4.3 Failure Reason: Not Accommodating Multiple Implementation Technologies 19
2.4.4 Failure Reason: Having a Central Standardization and Maintenance Authority 20
2.5 Comprehensive Data Standardization Meta Model 22

3.0 Metabase Demonstration 25

1.0 Frameworks

Error Statistic Summaries from Studies (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

Whitemarsh Knowledge Worker Framework						
Perspective	**Mission**	**Database Object**	**Business Information System**	**Business Event**	**Business Function**	**Organ-ization**
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 models and Identified Database objects**	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 models and Detailed Database Objects**	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 models**	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	**View data models**	Operating information systems	Start, stop, and messages	Detailed procedure based instructions	Daily activity executions, and assessments

Allocation of the GAO Information Technology Critical Success Factor Issues

Knowledge Worker Framework						
		Man-Machine Interface				
		Machine		Inter-face	Man	
Deliver-ables	Mission	Database Object	Business Infor-mation System	Business Event	Business Function	Organ-ization
Scope	5	2	3	1	3	4
Business	5	3	2	1	6	6
System	3	**2**	**2**	**1**	12	8
Tech-nology	1	**0**	**0**	**0**	8	6
Deploy-ment	0	**0**	**0**	**0**	5	5
Oper-ations	0	**0**	**0**	**0**	3	3
Col. Totals	**14**	**7**	**7**	**3**	**37**	**32**

Note: All numbers expressed as Percent allocations of errors to cells ...12 Gray cells are IT Cells

2.0 Data Standardization, The Key To Interoperability

- Typical Data Standardization "Problem"
- Testimonials????
- Semantics
- Common Reasons for Data Standardization Failures
- Data Standardization Components
- Data Standardization Work Plan
- Summary
- The Payoff

Just Why Is Data So Key? Because, Data Is Executed Policy

- If database is taken seriously its implementation cannot succeed without:
  - a standard data architecture,
  - an approach to accommodate diverse data naming, and
  - the ability to meld data across multiple database classes

- Data is executed policy.
- Data is the enterprise's persistent memory
- Data definitions are the technical representations of policy specifications.
- Policy executions, that is data, are the medium of business communication.

Key characteristics are:

- Business information systems that are a coherent union of the policy and then execution of the procedures that represent the accomplishment of the policy
- Consistent collection and/or modification of policy instances through the possible life cycle of the policy itself
- Consistent execution of policies when ever, where ever, and how ever deployed as the essence of the policy and the totality of its critical procedures are encapsulated within the database object class itself
- Minimized redundancy and consistent policy implementations across distributed environments as the database object class can be distributed through encapsulated strategies
- Comparable instances of deployed policies as their accomplishment is independent of hardware architectures and operating systems

2.1 Typical Data Standardization "Problem"

Seven different inventory systems:

2--Datacom/DB
1--IMS
1--DB/2
1--Pansophic (AS400) RPG/2 and ISAM
2--Oracle DBMS

For Inventory ALONE, there were:

- 1100 Columns, 30 different files. Hundreds of Billions of characters of existing data.
- Ten to twenty THOUSAND programs etc.
- All documentation is out of date. Systems have moved well past any existing documentation.
- World wide implementations (Europe, Canada, United States, Australia, Central & South America, Russia, and China)

Required Solution:

- One virtual system in which all names for the same concept appear the same, but none of the underlying systems are changed.
- All screens from all systems employ the same names for the same concepts
- All reports from all systems employ the same names in queries and on reports
- A single dictionary/repository in which unified semantics are stored
- On-the-critical path development environment at the LOGICAL level that has unified semantics regardless of physical implementation
- Automated documentation for the entire "virtual system." That is, requirements, specifications, user manuals, help desk support, etc.

2.2 Data Standardization is not just an abstract concept...

- A single data conversion and/or reformatting program is about 20 pages @ 50 lines per page. At 10 days of staff time per program (they're pretty simple), the cost would be about $5,000 per program.
- A single Government Agency currently spends $175,000,000 per year in such programs.
- That's 35,000 programs per year!
- For 175 million, you can:
  - ♦ Build 350 $500K houses
  - ♦ Educate 35,000 high-school students
  - ♦ Build 5% of an aircraft carrier
  - ♦ Buy 3,500 Mercedes

2.3 Semantics

- Semantics: Rules for meaning and usage
- Data Semantics: semantics for persistent data acquisition, storage, manipulation, and reporting
- Process Semantics: semantics for data transformations

2.4 Common Reasons for Data Standardization Failures

- A fundamentally flawed paradigm
- No Enterprise Wide Data Architectures
- Multiple implementation technologies
- Central standardization and maintenance authority

2.4.1 Failure Reason: Having a Fundamentally Flawed Paradigm

Traditional Paradigm: full name = prime word + modifier(s) + class word

Starting Point: TELEPHONE_NUMBER

Now, add modifiers:

- HOME_TELEPHONE_NUMBER
- OFFICE_TELEPHONE_NUMBER
- BOAT_TELEPHONE_NUMBER
- AIRPLANE_TELEPHONE_NUMBER

If the table was EMPLOYEE, then for employee, they were:

- EMPLOYEE_HOME_TELEPHONE_NUMBER
- EMPLOYEE_OFFICE_TELEPHONE_NUMBER
- EMPLOYEE_BOAT_TELEPHONE_NUMBER
- EMPLOYEE_AIRPLANE_TELEPHONE_NUMBER

And, when a new table came along, the standardization effort reoccurred.

For example, if the new table was CUSTOMER then standardization effort reoccurred for:

- CUSTOMER_HOME_TELEPHONE_NUMBER
- CUSTOMER_OFFICE_TELEPHONE_NUMBER
- CUSTOMER_BOAT_TELEPHONE_NUMBER
- CUSTOMER_AIRPLANE_TELEPHONE_NUMBER

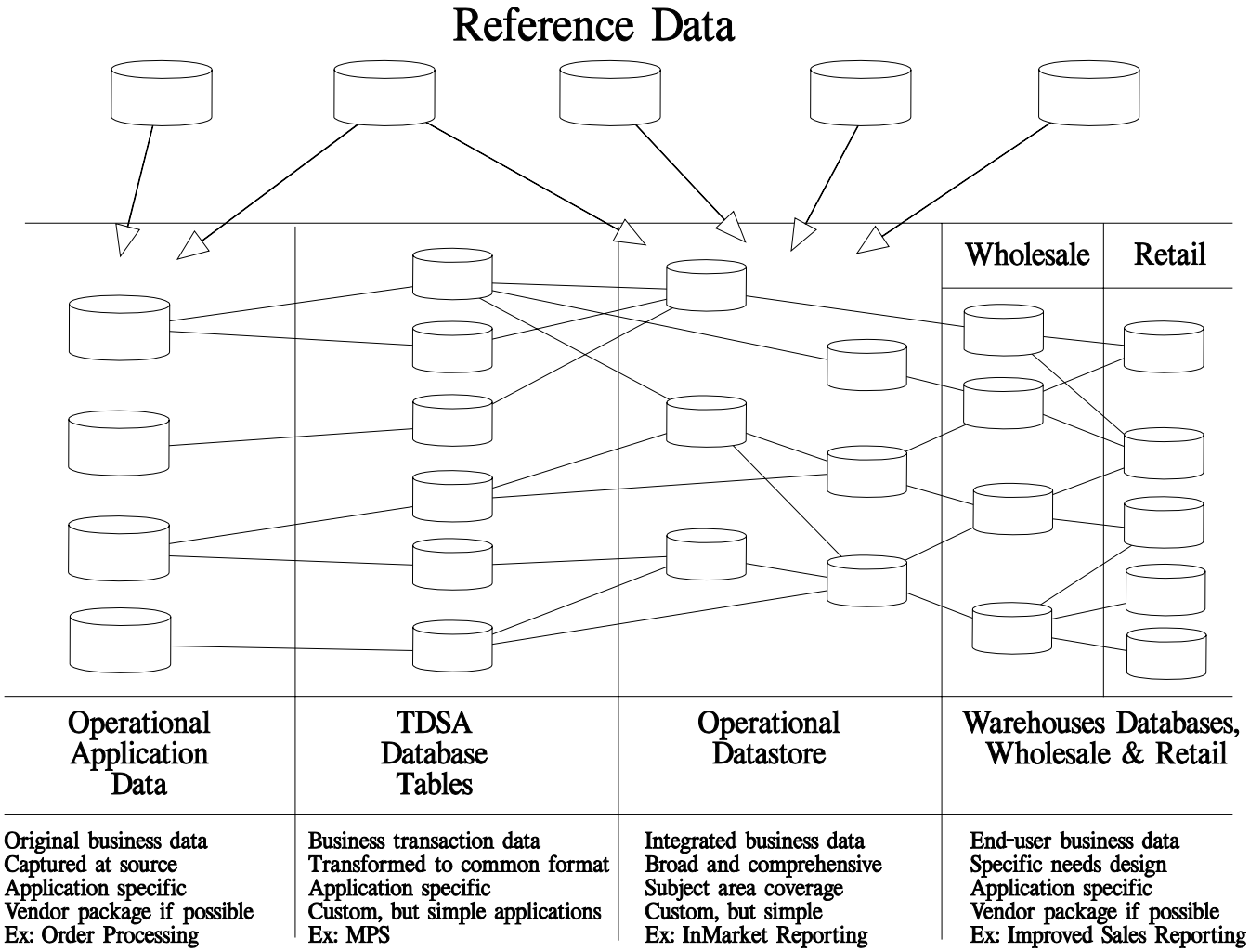
The Simple 3 Part Paradigm CREATES Problems

<prime_word>
<modifier>s
<class_word>

Examples:

EMPLOYEE SOCIAL SECURITY NUMBER (523-78-3872)	Number????
	Text??
	Code ????
	Identifier ???

2.4.2 Failure Reason: Not Dealing with Enterprise-Wide Data Architectures



Enterprise Wide Data Standardization Must Accommodate

Any effective data standardization effort must accommodate:

- Geography areas such as the world, regions, and countries
- Subject areas such as human resources, finance, marketing, manufacturing, sales and marketing, and research and development.
- Languages, where the same concept is represented differently
- Cultural/legal, where it is either traditional, required by law, or disallowed by law to collect and store certain information
- Finance systems that may be based on local laws and deal with sales accounting, finance such as accounts payable, receivable, and general ledger, when business is "booked," taxation, and payroll and compensation

Benefits from Enterprise Wide Data Standardization

- People have a better understanding of business and its state because data is accepted and understood by wider audience.
- Accelerates incorporation of new data and new uses of old-data because data sources are able to be quickly understood and are seen through standardized value discriminators.
- Reduces size of data by eliminating redundant data or data that is different merely for reason of style
- Frees up staff time to work on real business problem areas rather that ferreting out the same data hidden under different names.
- Increases the quality of decision making because data is valid, reliable, and represents discriminating facts about business activities

2.4.3 Failure Reason: Not Accommodating Multiple Implementation Technologies

- Different DBMS----IMS, Oracle, Datacom/DB, IDMS, Focus
- Different programming languages–Cobol, C, C++, PowerBuilder, Java
- Different data types, lengths, and value restrictions–Int, Char, VarChar, Decimal, Date, etc.

2.4.4 Failure Reason: Having a Central Standardization and Maintenance Authority

Starting point: Fully understood and ready to convey knowledge to "naming guru"

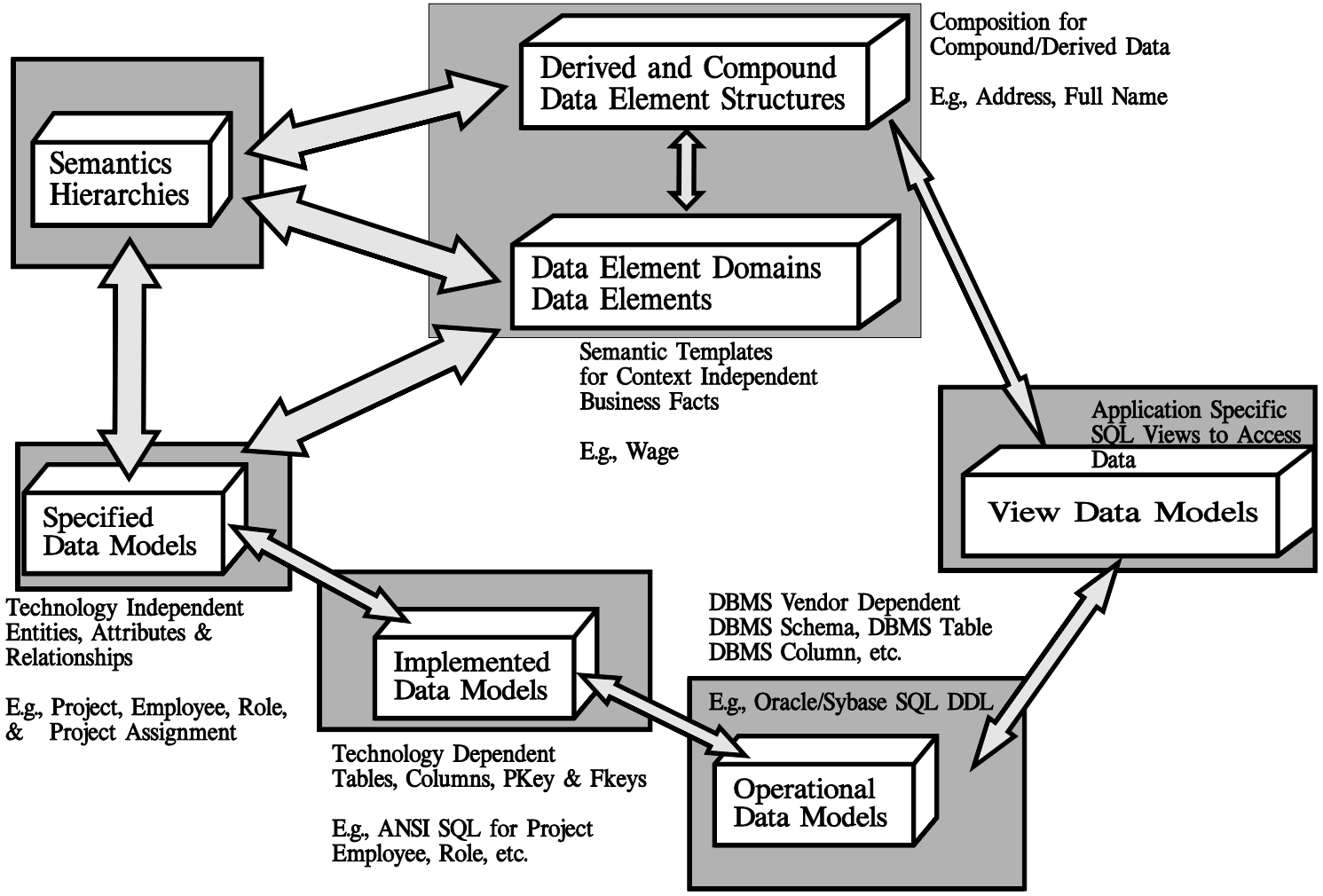
Effort for Data Standardization Per Year by a Centralized Staff					
Quantity of Projects	**Quantity of Names**	**Staff hours per project**	**Cost per project**	**Staff hours per year**	**Cost per year**
30 under development	500 names per project	240 hours	$18,000	7,200	$540,000
20 in production that need maintenance	900 names per year to maintain	23 hours of effort	$1,625	450	$32,500

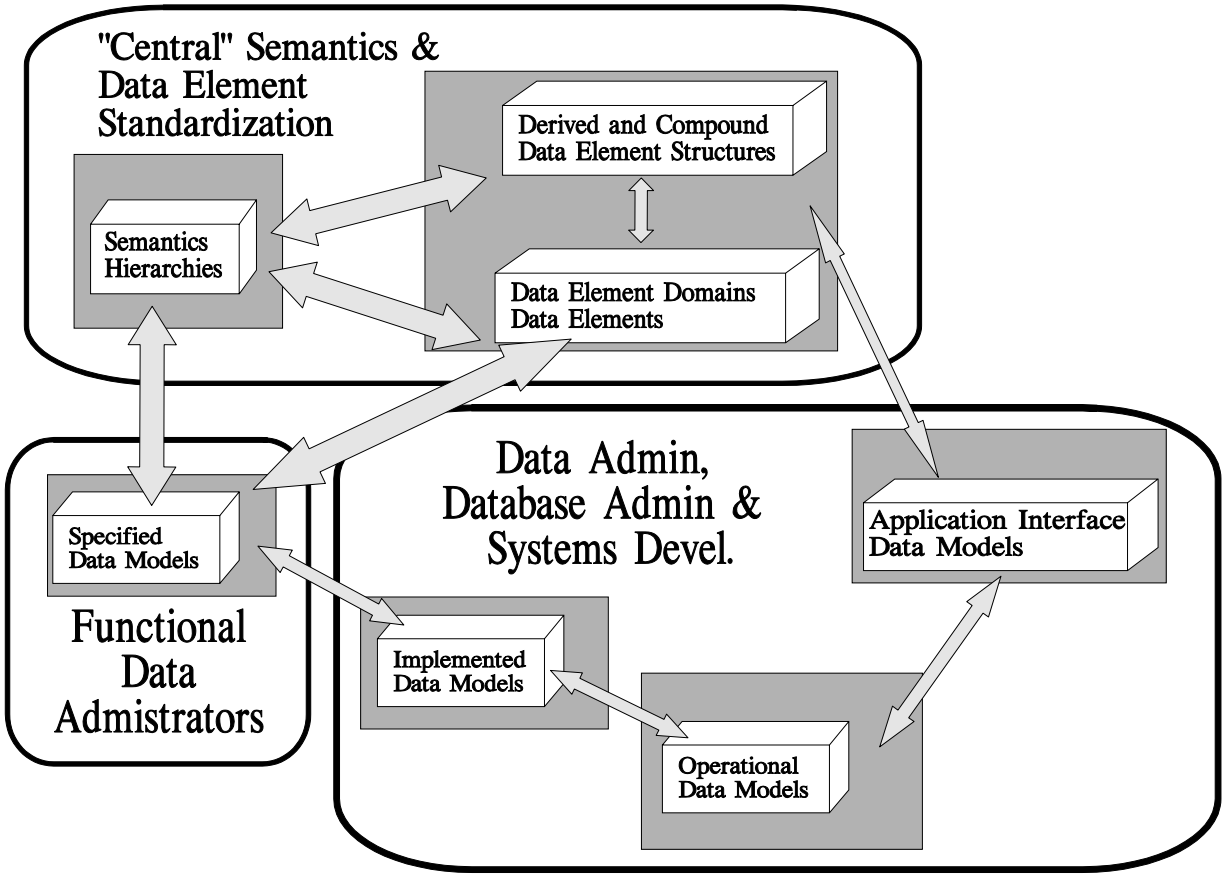
The ONLY correct "place" for a Centralized Standardization and Maintenance Authorityis at the Semantics and Data Elements

What would the cost be if there was a "system" that fully replaced the "naming guru?

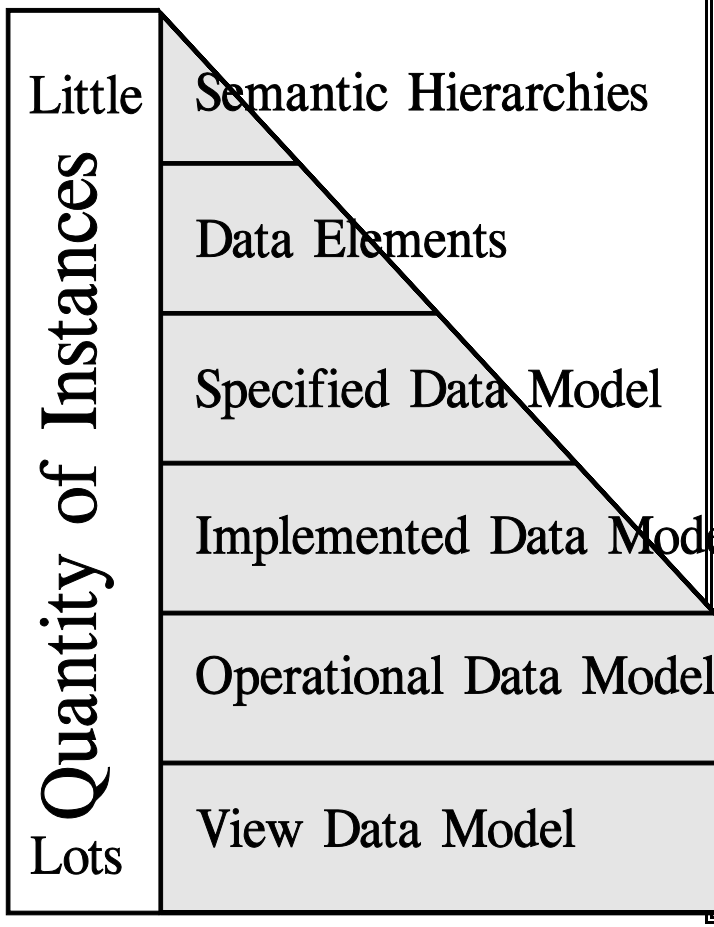
- Would there be any backlog?
- Would there be any under the table naming?
- Would there ever be a need for wavers from naming standards?

2.5 Comprehensive Data Standardization Meta Model



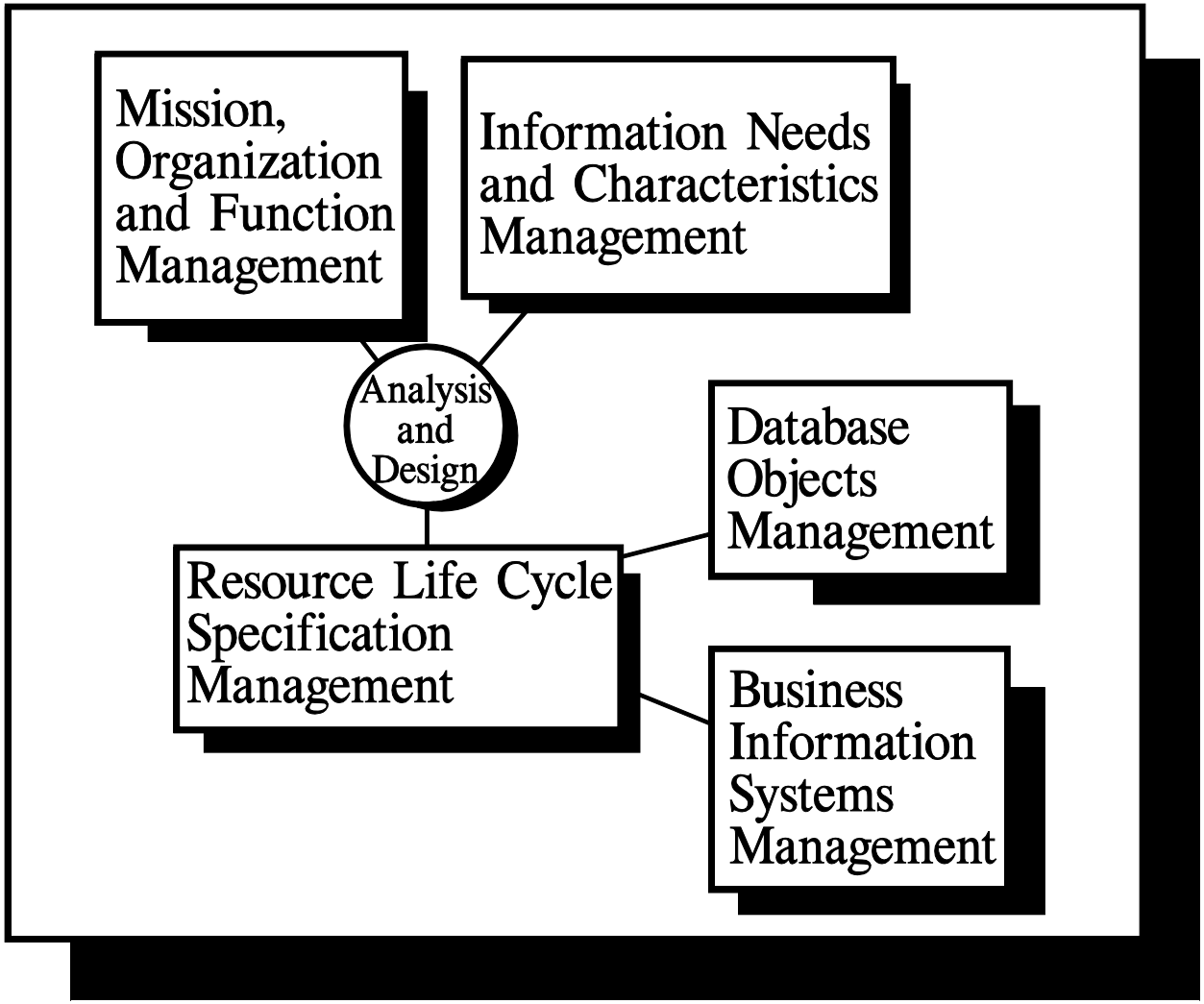


Distribution of Data Standardization Effort Across the Enterprise



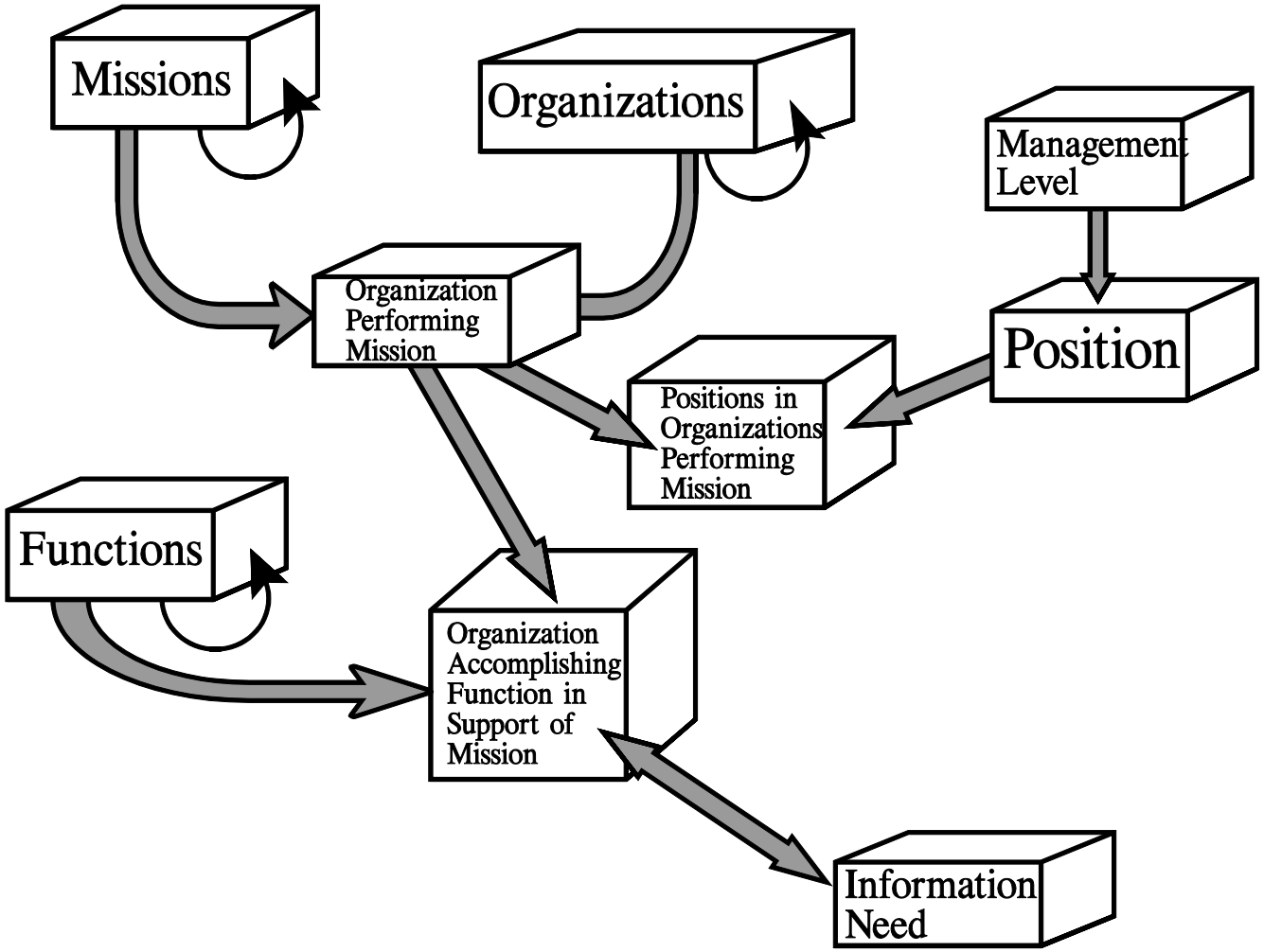
Key Metadata Models Essential for Data Standardization		
	Semantics	Absolutely and unconditionally, only one across the entire enterprise
	Data Elements	As few unique data elements within each database architecture class as possible
	Specified	As few as is possible for each different policy triple: {subject, entity, and attribute}
	Implemented	One for each database
	Operational	One for each deployed DBMS database
	Application Interface	A view for each database & program interface

3.0 Metabase Demonstration

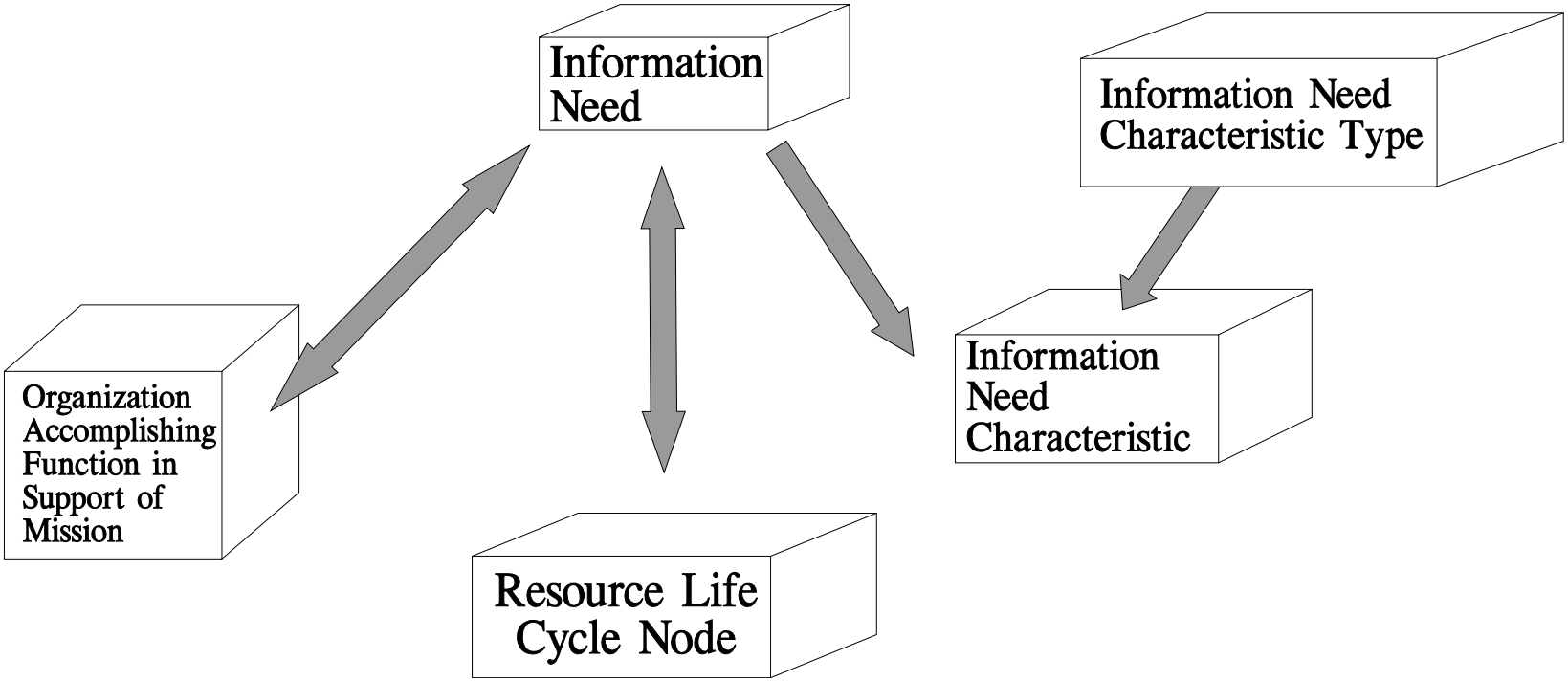


Metabase Software Module	Knowledge Worker Framework					
	Mission	Database Objects	Business Information System	Business	Business Function	Business Organization
Mission, Org. & Function	✔			✔	✔	✔
Information Needs			✔			
Resource Life Cycles		✔	✔			
Data Modeler: Data Elements		✔	✔			
Data Modeler: Specified		✔				
Data Modeler: Implemented		✔				
Data Modeler: Operational		✔				
Data Modeler: View		✔	✔			
Business Information Systems			✔			

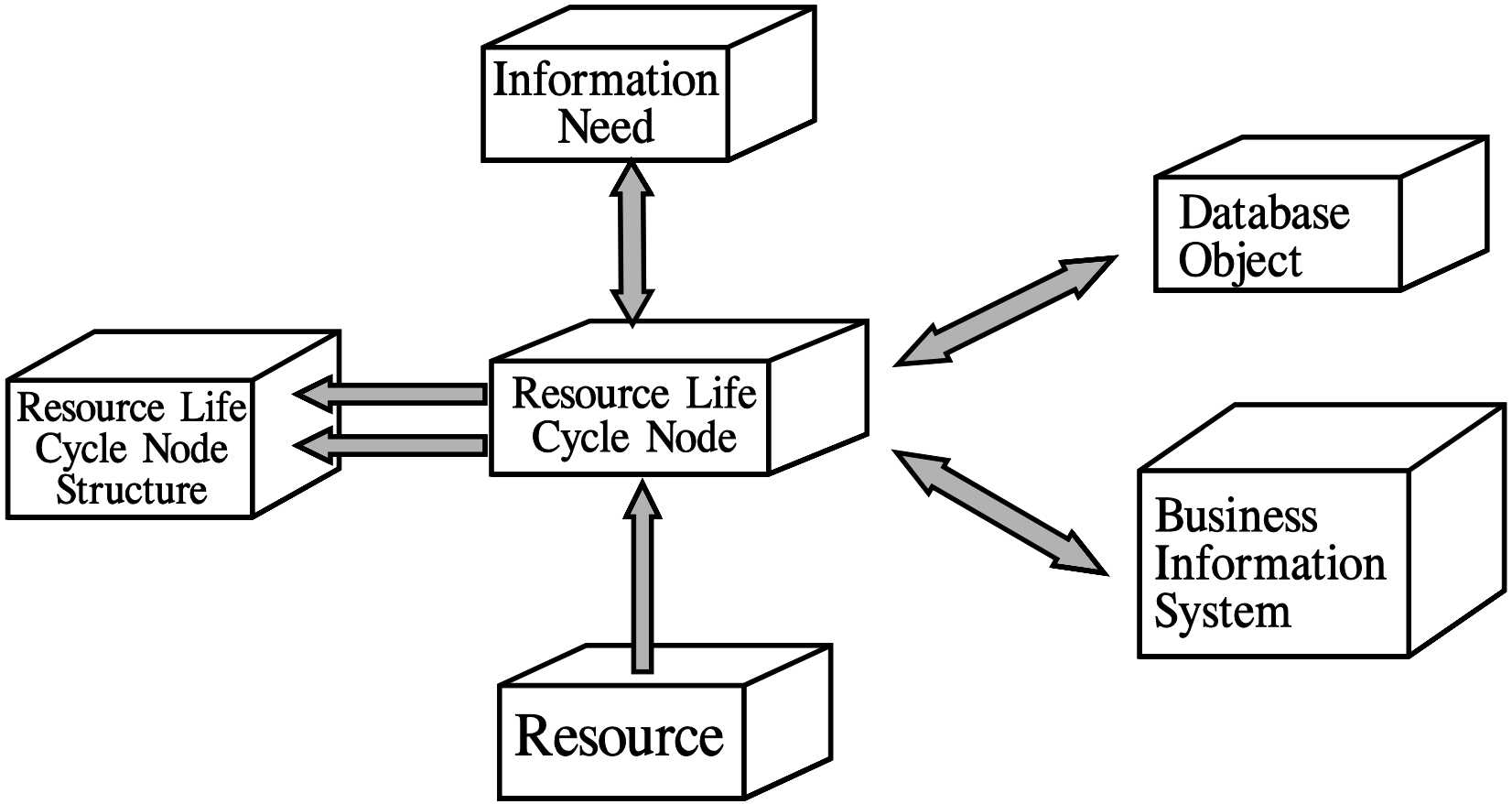
Mission, Organization, Functions and Positions



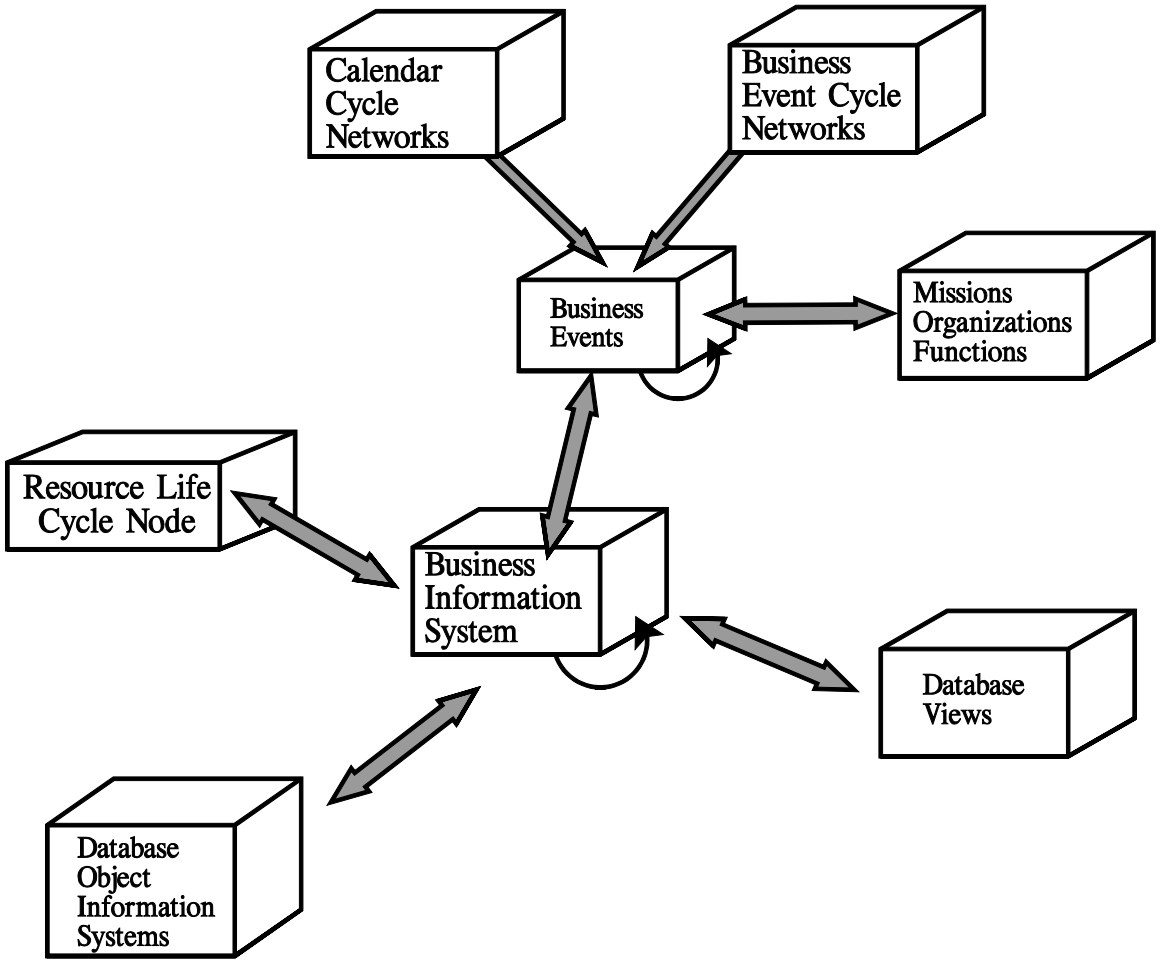
Information Needs Analysis



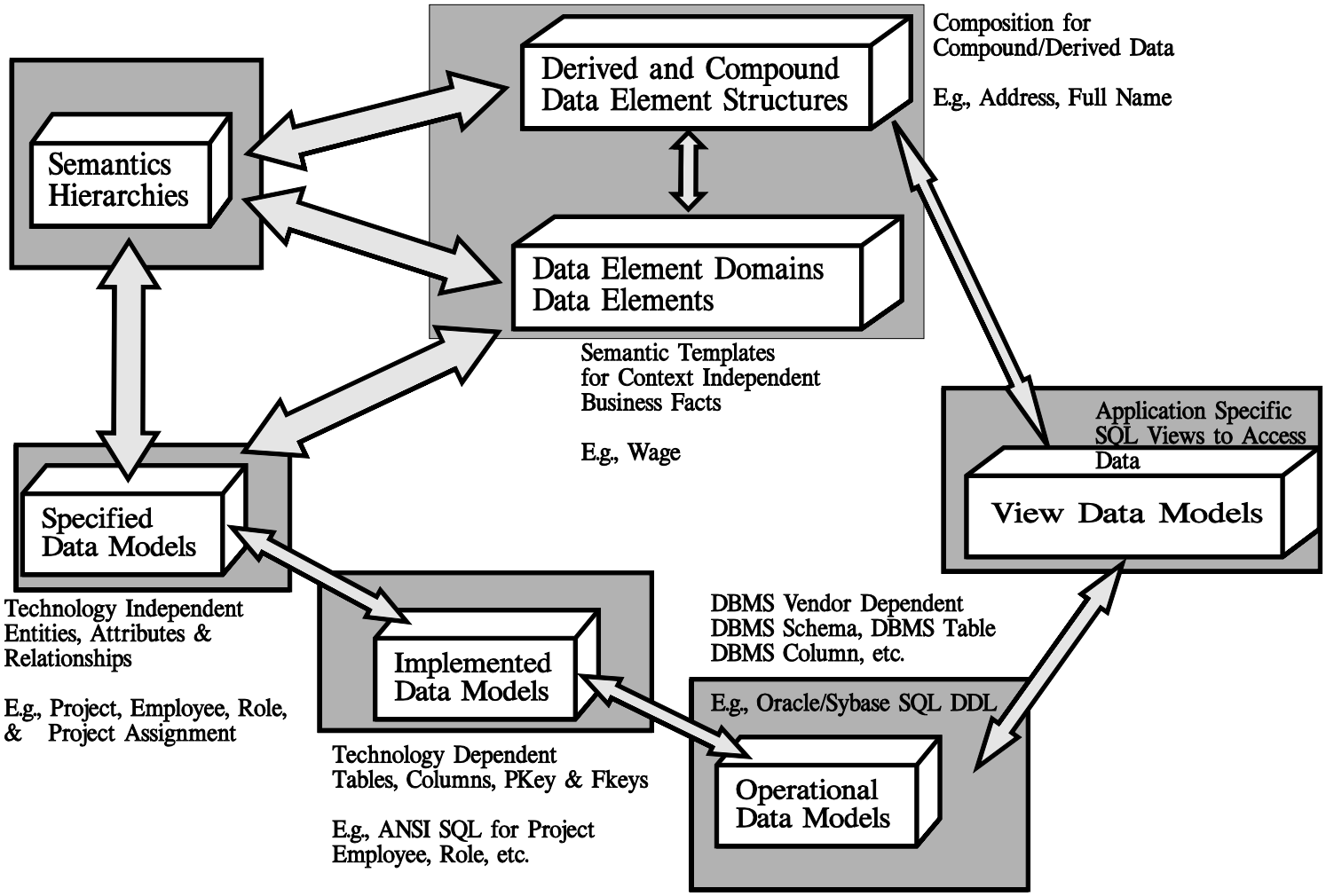
Resource Life Cycle Analysis



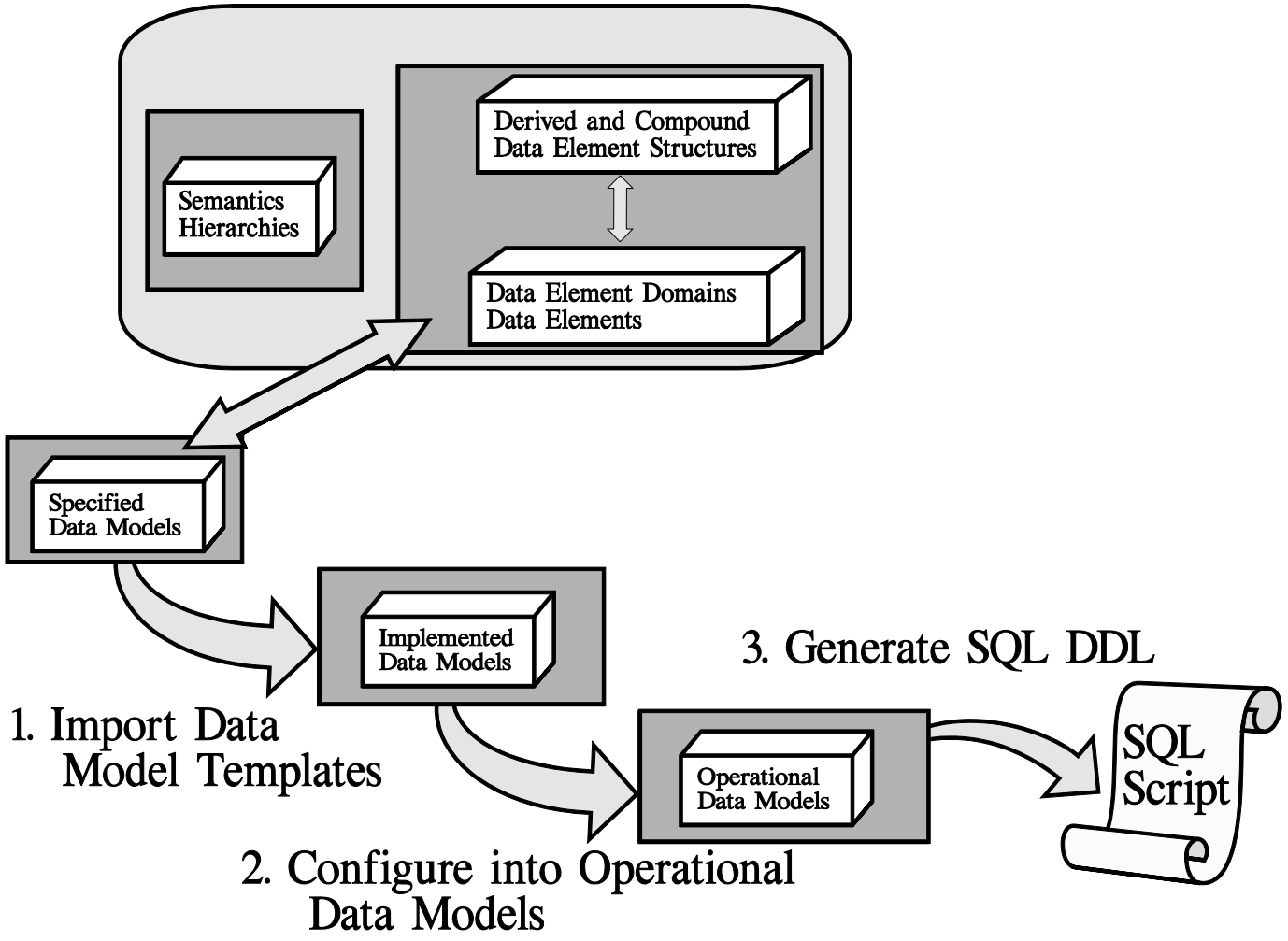
Business Information Systems



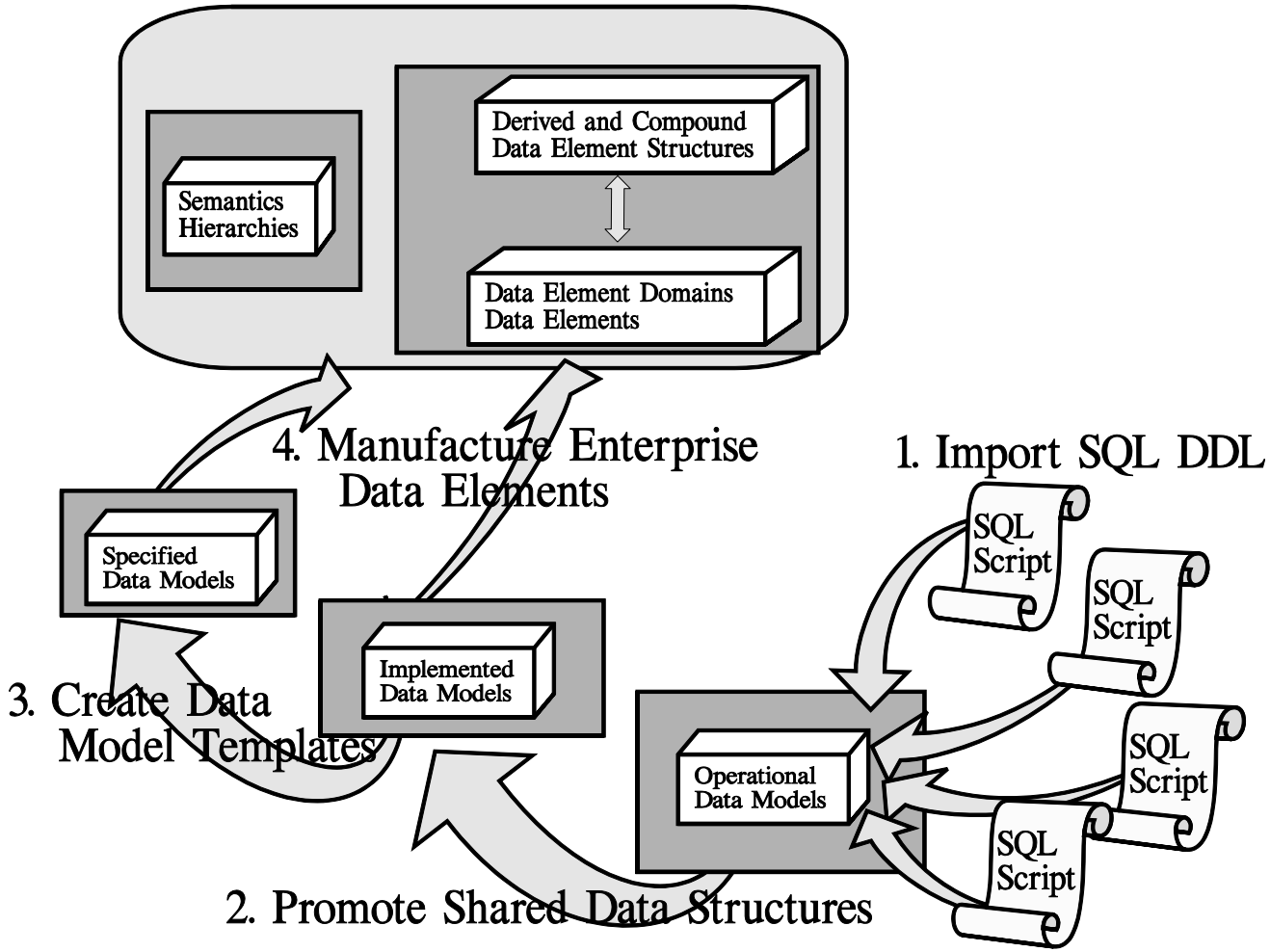
Data Elements, Specified, Implemented and Operational Data Models



Forward Engineering



Reverse Engineering



Metabase Demonstration Activities

- Sign-on
- Mission, Organization, Function, Person Assignment
- Information Needs Analysis
- Business Information Systems
- Resource Life Cycle Analysis
- Data Modeler
  - Data Elements
  - Specified Data Model
  - Implemented Data Model
  - Operational Data Model
  - View Model
- Database Objects

Metabase Demonstration	
Increased Productivity	Demonstrated increased productivity due to maximum reuse and interrelationships with other products. Highly engineered to minimize human effort and to eliminate error.
Increased Quality	Clearly shows consistency across deployments. Maximum reuse of commonly defined components across the Knowledge Worker Framework
Decreased Risk	Nothing is missed and nothing extra is accomplished. Minimum essential and maximum accomplished are the same.
Decreased Cost	Employable by multiple projects and staff within the enterprise. Change impact is easily deduced and any required changes are able to be planned, accomplished and tracked.