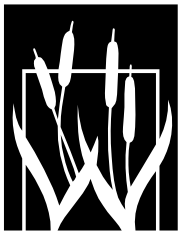


Whitemarsh
Information Systems Corporation

Achieving Data Standardization in a World-Wide Heterogeneous Database Environment

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

Table of Contents

1.0 The Issue

- 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

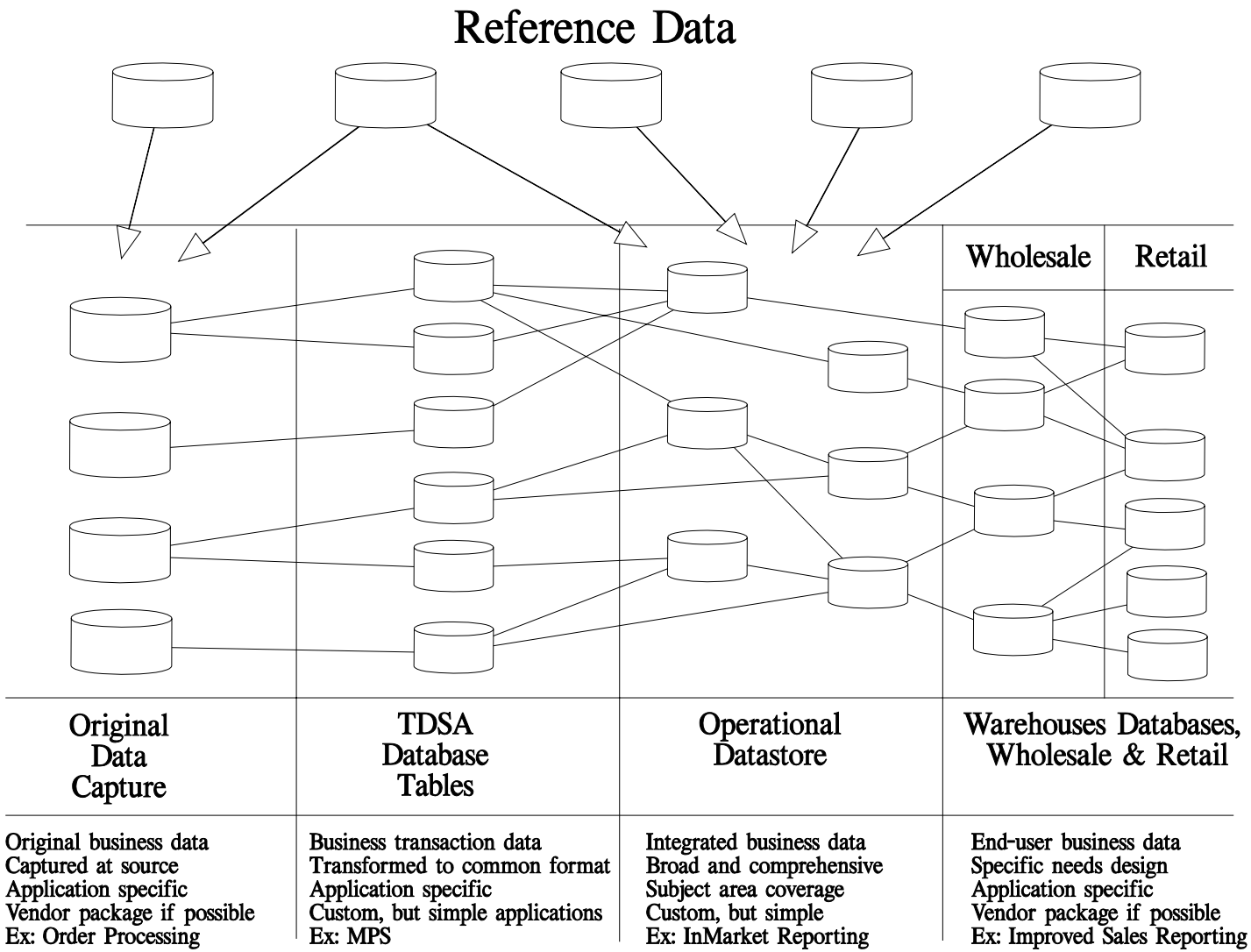
- Data is executed policy.

- Data is the enterprise's persistent memory

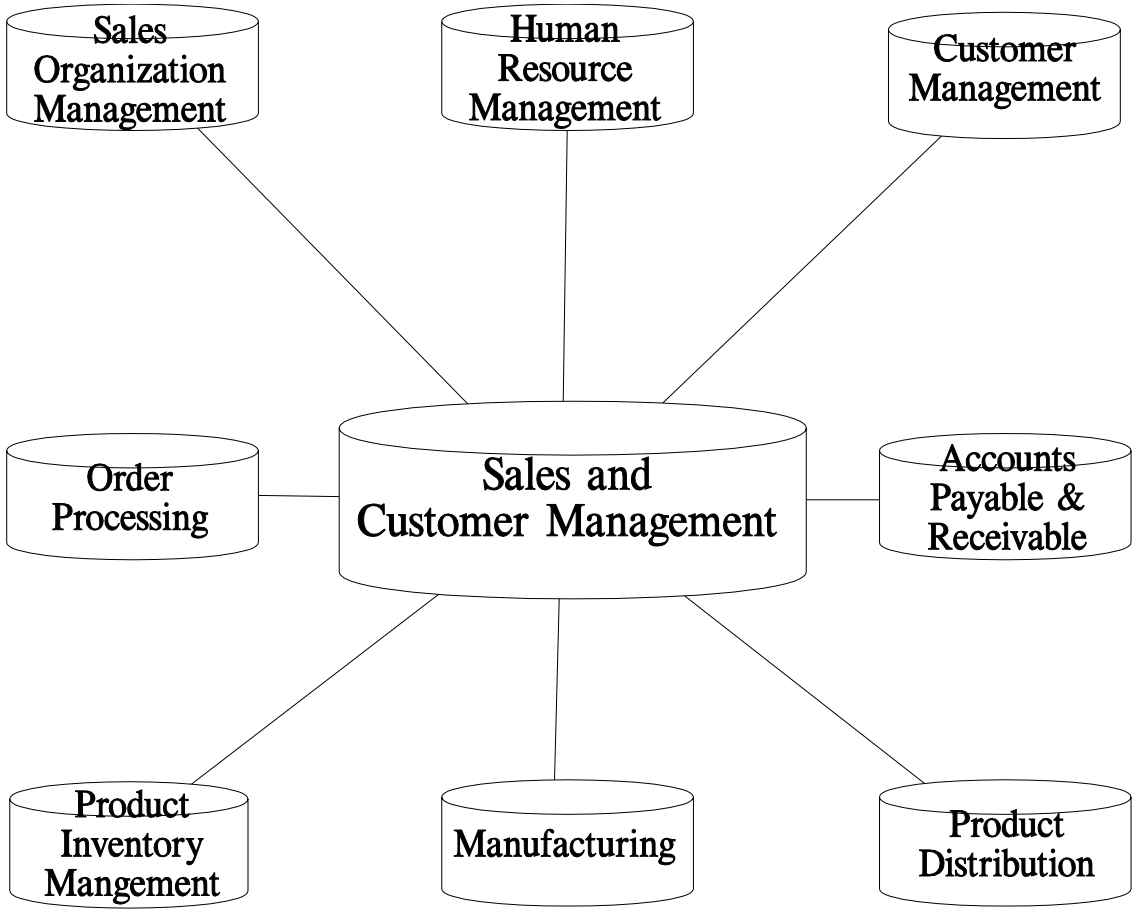
- Data definitions are the technical representations of policy specifications.

- Policy executions, that is data, are the medium of business communication.

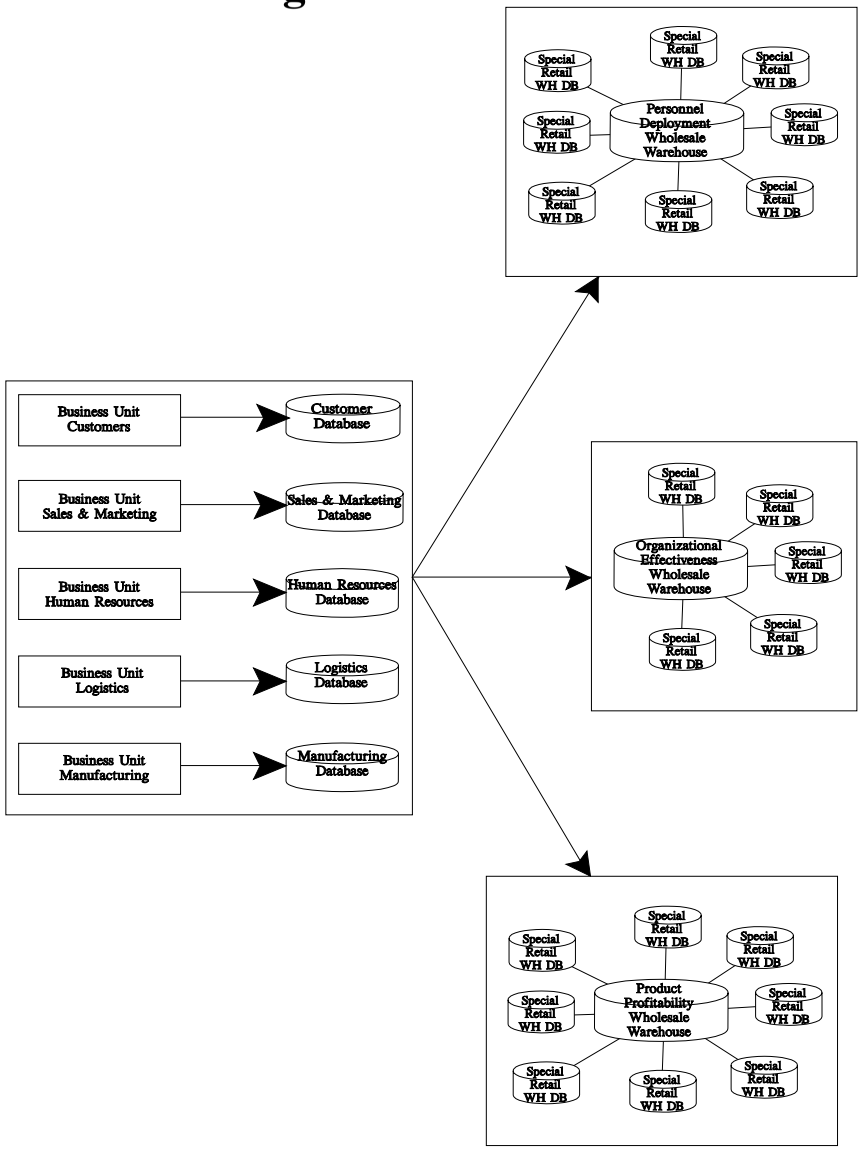
2.0 Data Architecture Classes



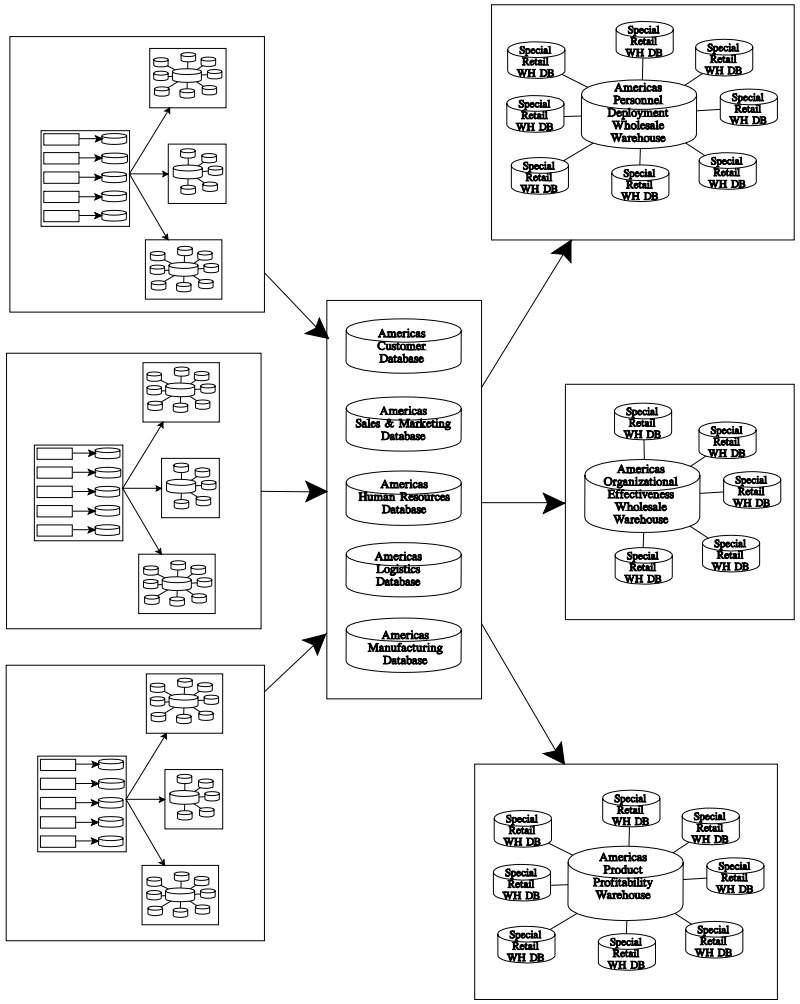
2.1 Typical Transaction Database and Subject Area Database Configuration



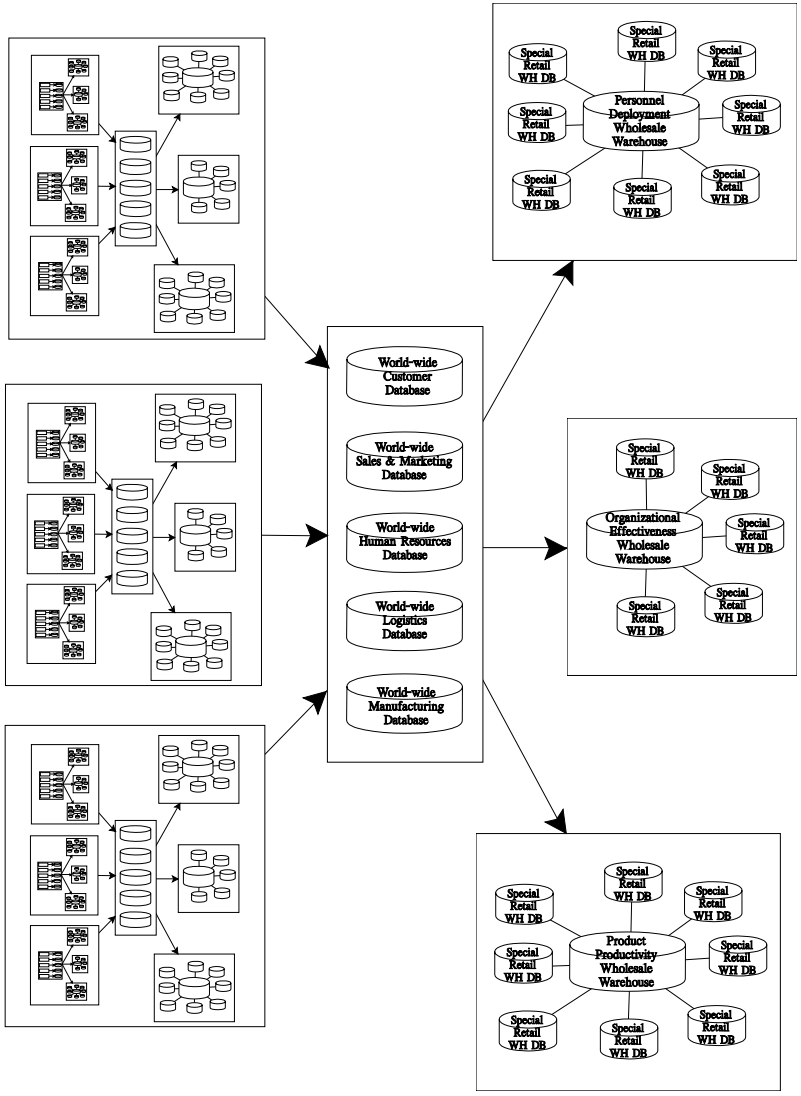
2.2 Typical Data Warehouse Configuration



2.3 Typical Country-wide Multi-site On Line Transaction Processing (OLTP), and On-Line Analysis Processing (OLAP) Database Configuration



2.4 Typical World-wide Multi-Level Sites of OLTP and OLAP Database Configuration



2.5 Characteristics of Each Data Architecture Class

Data Architecture Persistent Data Classifications and Characteristics				
Persistent Data Classification	**Persistent Data Characteristics**	**Process Characteristics**	**User Considerations**	**Technical Considerations**
Original Data Capture data	Detailed atomic data Accurate as of the last update Well defined, long lasting database designs Normalized database designs Uses reference data No invalid data updates allowed	Tuned for transaction capture, storage and update Application oriented Transaction driven Processing supported by well known data integrity and business processing rules Understands, creates, and maintains TDSA databases through operational application system extract and TDSA loading software. Source data for TDSA	Original data source entry personnel High availability Supports day-to-day operations	Amount of data for processing is small Multiple vendor packages Package specific May or may not be controlled by SQL DBMS

Data Architecture Persistent Data Classifications and Characteristics				
Persistent Data Classification	**Persistent Data Characteristics**	**Process Characteristics**	**User Considerations**	**Technical Considerations**
TDSA: Transaction Data Staging Area	Transient data/short lived Foundation data source for all operational data systems Mars-wide standard semantics Package independent database designs Denormalized Full business transaction Does not use reference data	Accepts, stores, and then pushes forward function Only refreshed with changes from previous version Operations application data system daily update event driven Translation and transformation	Users cannot access	Multiple platforms Interface monitoring Applications insulation SQL DBMS controlled

Data Architecture Persistent Data Classifications and Characteristics				
Persistent Data Classification	**Persistent Data Characteristics**	**Process Characteristics**	**User Considerations**	**Technical Considerations**
ODS: Operational Data Store **"Subject Area Data Store"**	Detail level data May be lightly summarized Current or nearly current Rolling histories Broad subject area database scope Normalized database designs Redundant data from across enterprise May contain derived data from "outside" Uses reference data	Updated daily via TDSA data transaction files Accepts and stores data from TDSA Supports comprehensive reporting and generalized ad hoc query	End-user detailed level analysis Used for up to the minute decisions Used for detailed decision making	Requires fast response time Large volume SQL DBMS controlled

Data Architecture Persistent Data Classifications and Characteristics				
Persistent Data Classification	**Persistent Data Characteristics**	**Process Characteristics**	**User Considerations**	**Technical Considerations**
Warehouse: Wholesale	Summarized and some detail Rolling Histories Load/replace, no end-user update Enterprise-wide standard semantics Narrow/subset of one or more subject areas Redundant data from across enterprise May contain internal derived data Reference data fully embedded May receive and/or	No end-user updating Regular, periodic updates Supports standardized, on-demand reports Supports general complex business data analyses such as trends and forecasting Views data from multiple subject areas	Supports managerial community Used for broad direction and positioning Used to formulate and assess long term decisions	Availability not on business' critical path User workstation access Large data volumes per query High processing power required SQL DBMS controlled

Data Architecture Persistent Data Classifications and Characteristics				
Persistent Data Classification	**Persistent Data Characteristics**	**Process Characteristics**	**User Considerations**	**Technical Considerations**
Warehouse: Retail	Light to highly summarized and some detail Rolling histories Load/replace, no end-user update Enterprise-wide standard semantics Denormalized and highly designed to specifically favor one or more reporting formats Redundant data from across enterprise May contain internal derived data Reference data fully embedded	Availability not on business' critical path Regular, periodic updates Highly designed, end-user on-demand reports Supports very specific simple to complex business data analyses Views data from multiple subject areas	Supports managerial community Cannot update Used for direction and positioning Used for long term decision making Specific reporting need	Relaxed availability User workstations Large volume High processing power SQL DBMS controlled

Data Architecture Persistent Data Classifications and Characteristics				
Persistent Data Classification	**Persistent Data Characteristics**	**Process Characteristics**	**User Considerations**	**Technical Considerations**
Reference Data	Durable codes and long value alternatives with policy definitions and full descriptions Mars-wide standard semantics Source of all valid and invalid values including alternatives for different countries and languages Multiple group data field constructors suitable for different countries and languages Definitive source for multi-use data in all other databases	Simple updates Update mappings required for reference data value migration	Needed by all levels in the organization Used by all systems Enables understanding and conversion of historical data	Supports the concept of single source Integration with all data store types

2.5 Benefits from Distinct Data Architecture Classes

- 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

3.0 Semantics

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

3.1 Data Semantics Problems

When the same concept is known under two different semantic representations. Data semantics irregularities are most commonly evidenced through differences in:

- data names
- data types
- data lengths/precisions
- data structures

Examples:

Name Problems: SSN vs Social Security Number
Type Problems: SSN INT(9) vs SSN Char (11)
Precision Problems: ANNUAL_OVERTIME_HOURS (Decimal)

Structure Problems:
Inventory Quantity January, ..., Inventory Quantity December
Vs
Inventory Year, Inventory Month, Inventory Quantity

3.2 Process Semantics

Process semantics are the rules that govern data transformation. Typical problems include:

- Improper specification
- Wrong placement leading to multiple, but different executions
- Improper maintenance and evolution

Examples:

Improper Specification: Computing an average but not considering null values to reduce count

Encoding different data quality process in different programs. Given: Sex Char(1),

Sex: Y|N|U
Sex: M|F
Sex: 0|1|2

Updating 98% of all the instances of someone's address. That is, not having "the golden source" and then reference data replication.

3.3 Testimonials????

- 50% of all software costs are attributable to error corrections–The U.S. DoD
- 60% of all corporate IS budgets are devoted to correction of errors–Software Quality by Mordecai Ben-Menachem and Gary Marliss
- Software maintenance time: 47% analysis, 28% testing & debugging, 19% coding, and 6% documentation--errors–Software Quality by Mordecai Ben-Menachem and Gary Marliss
- 700,000 Americans were shorted $850,000,000 in Social Security payments due to software error.–U.S. Social Security Administration
- 3 times more errors are introduced during maintenance than in original development.

Data Standardization Is Not Just an Abstract Concept...

1. 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 (1995 study).

2. The U.S. Government spends $ 4.175 Billion per year in such programs (1995 study).

3. That's 835,000 programs per year!

4. For $ 4.175 Billion, you can:

   - Build 167,000 houses
   - Educate 417,000 high-school students
   - Buy 83,500 Mercedes

4.0 Common Reasons for Data Standardization Failures

- A fundamentally flawed data standardization model
- No accommodation for enterprise wide data architectures
- Multiple implementation technologies
- Central standardization and maintenance authority

4.1 Failure Reason: Having A Fundamentally Flawed Data Standardization Model

The Agency REQUIRED the PRIME word to prefix every "data element," thus, every "data element" was really a column.

Every staff hour spent standardization columns is a 4x waste of time:

- First was the effort to standardize the column.
- Second was the time to recognize the error and have the courage to admit to such a callosal waste of effort.
- Third was the effort to find all the redundant definitions.
- Fourth was the effort to delete the definitions and to create the single definition as well as all the mappings from the defined once to the used many times locations.

Standardizing columns is a non-productive, ever-widening effort

- Focus was Too Low: The agency's standardization effort concentrated on a level of detail (columns) that is ever expanding, thus prevented completion of the effort.
- Standardization was overly focused on *names.*
- Critical components requiring standardization were being ignored.
- The current effort attempted standardization without critical context and subject matter materials.

Standardizing columns is a non-productive, ever-widening effort(cont.)

- The agency's repository tool, while adequate for definition and maintenance of standard columns, was unsuited to the overall task because:
  - ♦ It concentrated on columns, which was a level of detail that was ever expanding,
  - ♦ It lacked critical analysis and synthesis capabilities,
  - ♦ Its meta-model was insufficient for a comprehensive and correct data standardization effort.
- The existing procedures excluded critical areas: This caused intense standardization of too small an area of the overall data standardization problem, almost to the exclusion of other critical areas.

4.1.1 The Focus of Standardization Was Too Low

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:

- CUSTOMER_HOME_TELEPHONE_NUMBER
- CUSTOMER_OFFICE_TELEPHONE_NUMBER
- CUSTOMER_BOAT_TELEPHONE_NUMBER
- CUSTOMER_AIRPLANE_TELEPHONE_NUMBER

4.1.2 Standardization Was Too Focused on *Names*

Naming standards that focus on recording the codification of data semantics into the data's name commonly obscure the needs of the business and the common business meaning of the data.

- When the data element TELEPHONE NUMBER is properly named standard column, it becomes:

  EMPLOYEE_TELEPHONE_IDENTIFIER

- NUMBER is missing because it is a class word. IDENTIFIER must be used since telephone number is really a form of a unique address of a telephonic instrument.

4.1.2 Standardization Was Too Focused on *Names* (cont.)

- SOCIAL SECURITY NUMBER, is not really a number. In fact it is a code as well as an identifier thus its name should really be:

  EMPLOYEE_SOCIAL_SECURITY_IDENTIFER_CODE

but that would violate the *rules* because only one class word was allowed. Thus it had to be either:

EMPLOYEE_SOCIAL_SECURITY_IDENTIFER

EMPLOYEE_SOCIAL_SECURITY_CODE

4.1.3 Critical Standardization Efforts Were Ignored

No standardization of:

- Derived data, e.g., FINAL_MONTHLY_BALANCE
- Compound, e.g., FEDERAL_STOCK_NUMBER

Not fully defining compound data elements,

- discards very important tradition,
- invites malformed update processing, and
- invites its subsequent reintroduction when it cannot be "discovered" among all the many other primitive and atomic data elements.

4.1.4 Critical Context and Subject Matter Materials Were Missing

Because of a very controlled data standardization effort supported by complete documents,

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

4.1.5 Repository Tool Was Unsuited for the Task

- Flawed meta model
- No connectivity to production databases and systems
- Pull on accidental discovery vs Mandatory Use Push data semantics
- Impossible metadata cleansing due to ever widening effort

4.1.6 The Existing Procedures Excluded Critical Areas

- Standardization result was not woven into the fabric of systems development.
- Standard names and definitions immediately began to unravel
- Maintenance was never incorporated

4.2 Failure Reason: No Accommodation for Enterprise Wide Data Architectures

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

4.3 Failure Reason: Not Accommodating Multiple Implementation Technologies

- Different DBMS
- Different programming languages
- Different data types, lengths, and value restrictions

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

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?

4.5 Standardization Problem Summary

To be successful, a data standardization effort must:

- Properly focus on the data concepts such that the lowest level column is easily and automatically named.
- Accommodate enterprise wide data architectures such that data contexts are immediately apparent without compromising data identification, selection, and melding.
- Accommodate multiple implementation technologies such that regardless of the rules, a data concept are immediately obvious and relatable to all other data that espouse the same concept.
- Allow front-line project staff to create and maintain their own names under the guidance and work enhancing tools and techniques of a central standardization and maintenance authority

5.0 Data Standardization Parts

- **Business Policy Semantics**: the formulation of the business policies that must be represented in both data and process based transformations of the data. (Section 6)

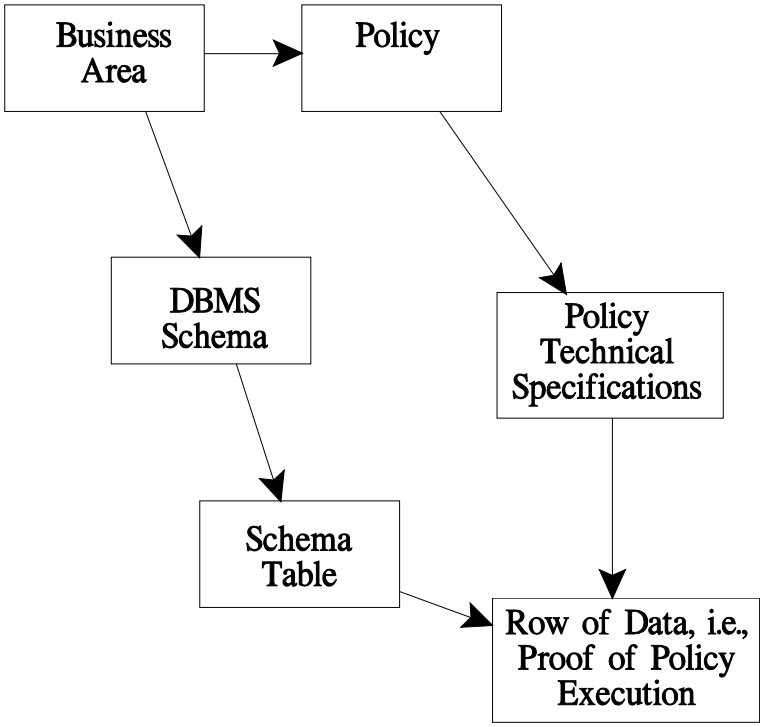
- **Data Semantics**: The expression of the business policy semantics in metadata to support the efficient and effective definition, deployment and maintenance of data and processes throughout a world-wide, heterogeneous database environment. (Section 7)

- **Data Value Sets**: The expression of executed business policy as values according to the accepted data semantics. (Section 8)

- **Comprehensive Work Plan**: The human processes necessary to achieve (i.e., define, implement, and sustain) data standardization throughout an enterprise in an efficient and effective manner. (Section 9).

6.0 Business Policy Exposition

- Policy is represented through policy technical specifications
- Policy technical specifications are manifest through rows of data
- Schema tables are represented through rows of data



Business policy, that is, the details that support data meaning and usage convey the fundamental business reason for collecting, storing and maintaining data. For example,

Marital Status is defined as the identification of the mutually exclusive states regarding a person's marital state. The two states are: single and married.

The states (single and married) must be based on corporate policy as the official states to record a person's marital status.

7.0 Data Semantics

7.1 Areas Embraced

Data Semantics Component	Specification	Implementation
Policy homogeneous business segment	Subject Area	Schema
Well defined unit of business policy	Database Object	SQL Table
Precise set of data within a business policy	Entity	SQL Table
Discrete evidence of policy adherence	Attribute	Column
Unique discriminator of policy instance	Primary Key	Primary Key
Relationship mechanism between policy instances	Foreign Key	Foreign Key
Business rule within a business fact	Type of Data Integrity Rule	Column Constraint
Business rule for whole business policy	Type of Data Integrity Rule	Table Constraint, Assertion
Business action consequence when a policy is accomplished or violated	Type of Data Integrity Rule	Trigger or Stored Procedure

7.2 Two Alternatives

- Standard semantics (data names, data types, data lengths, and data structures) for entities & attributes, and tables & columns. That is, where ever the data semantic concepts are the same they must be represented the same way.

- Semantics mappings from standard, non-redundant, technology and context independent concepts (data names, data types, data lengths, and data structures) to technology and context dependent specified and implemented uses of those concepts.

The first approach is **not** possible because:

- The accepted data semantics now in place force all names to be different. Thus concepts that are the same cannot be found because they are at the very least named differently.

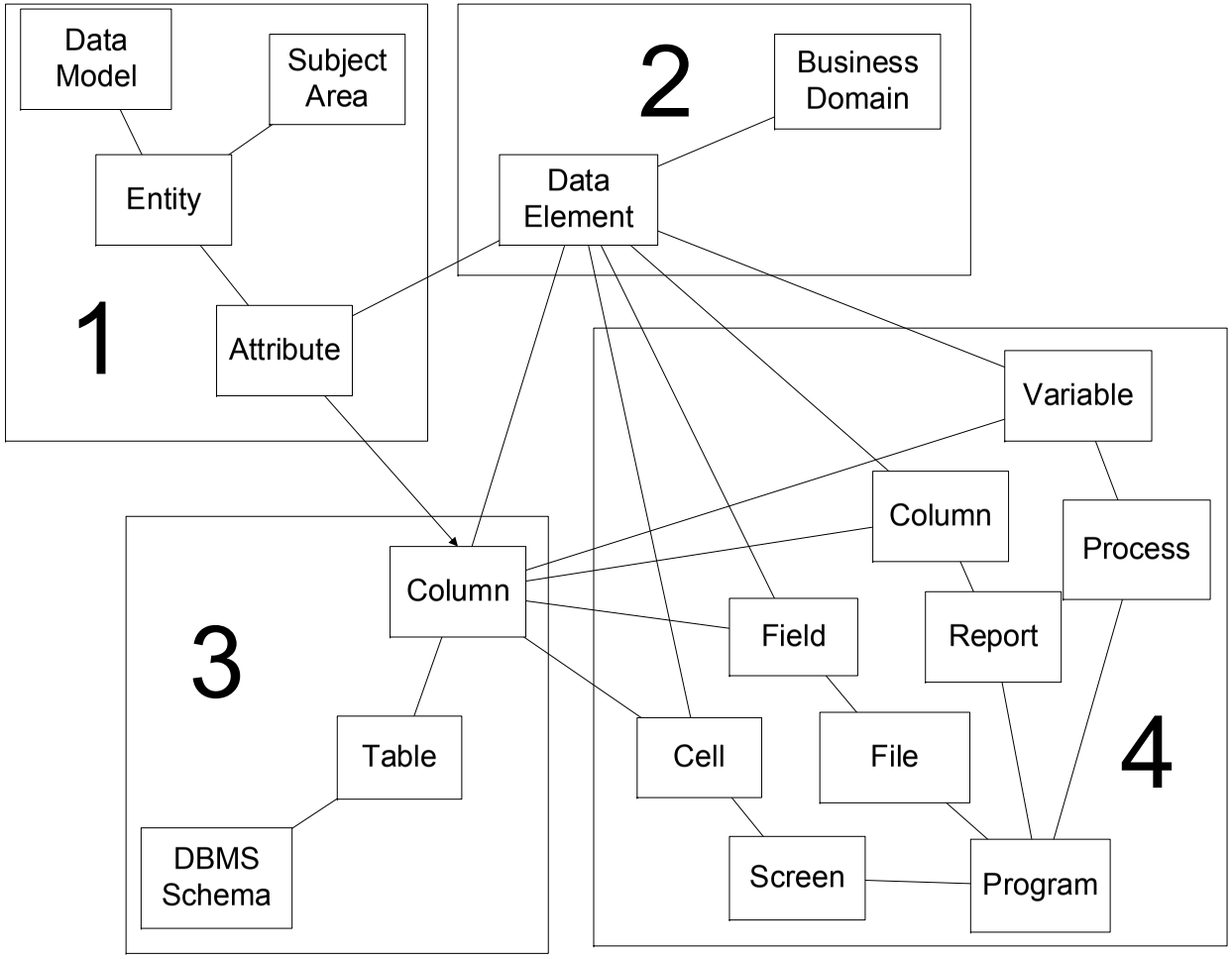
- It requires abandonment of all COTS procurements because different semantics across package within overlapping database domains. For example, numerous COTS application packages all involve semantically different personnel information that addresses the same or essentially similar concepts.

- It requires a complete rewrite of all legacy applications from their non-standard semantics to a set of standard semantics. And the continuous rewrite of these legacy systems every time the semantics standard changes.

The semantics mappings approach requires addressing:

- Subject Area/Schema, which is a policy homogeneous business segment
- Database Object, which is a well defined unit of business policy
- Entity/Table, which is a precise set of data within a business policy
- Attribute/Column, which is a discrete evidence of policy adherence
- Primary Key, which is a unique discriminator of policy instance
- Foreign Key, which is a relationship mechanism between policy instances
- Attribute Data Integrity Rule or Column Constraint, which is a specific business fact rule
- Entity Data Integrity Rule or Table Constraint, which is a specific assertion business rule that affects a whole business policy instance
- State Data Integrity Rule, Trigger or Stored Procedure, which is a business action consequence when a policy based state is accomplished or violated

7.3 Data Semantics Meta Model



7.4 Data Semantics Components

Meta Attributes for Data Elements and Context Dependent Business Facts		
Meta Attribute or Semantic Part	**Business Data Element**	**Entity Attribute, File Field, Screen Cell etc**
Business Domain (See 7.7)	Yes	No
Common Business Name (See 7.8)	Yes	Yes
Prime word	No	Yes
Modifier subclasses (See 7.9)	No	Yes
Class word subclasses (See 7.10)	No	Yes
Generated Policy Basis Description	Yes	Yes
Computer Data Type	No	Yes
Data Structure	Yes	Yes
Required Uniqueness	No	Yes
Relationship Function	No	Yes

7.5 Old 3 Part Paradigm Not Viable

<prime_word> <modifier>s <class_word>

Examples:

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

Example: TELEPHONE_NUMBER

HOME_TELEPHONE_NUMBER
OFFICE_TELEPHONE_NUMBER
BOAT_TELEPHONE_NUMBER
AIRPLANE_TELEPHONE_NUMBER

If the table is EMPLOYEE, then you have:

EMPLOYEE_HOME_TELEPHONE_NUMBER
EMPLOYEE_OFFICE_TELEPHONE_NUMBER
EMPLOYEE_BOAT_TELEPHONE_NUMBER
EMPLOYEE_AIRPLANE_TELEPHONE_NUMBER

Then additional sets for:

Customer
Dependent
Supplier
etc.

For each set you need names, definitions, rules, programs, tests, etc.

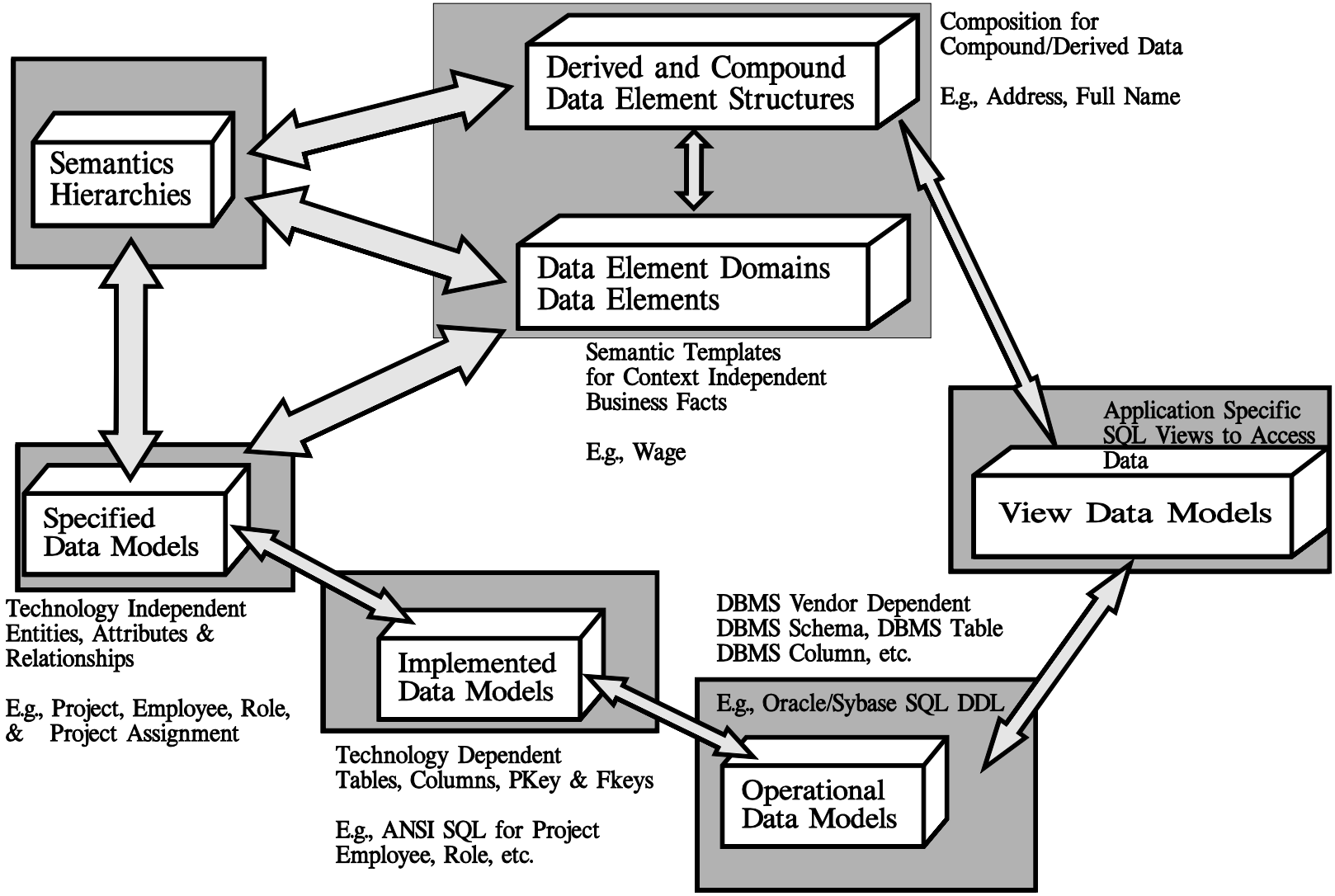
The problems with the old three part name scheme are:

- having prime a word,
- not having a common business name,
- heterogeneous modifiers and class words, and
- only allowing one class word.

Telephone number is really telephone number, no matter how, when, or why used

Social Security Number is really that Id thing used by the US Government. It's not a number, a code, or anything else!

7.6 A New Paradigm & Meta Model Basis is Required



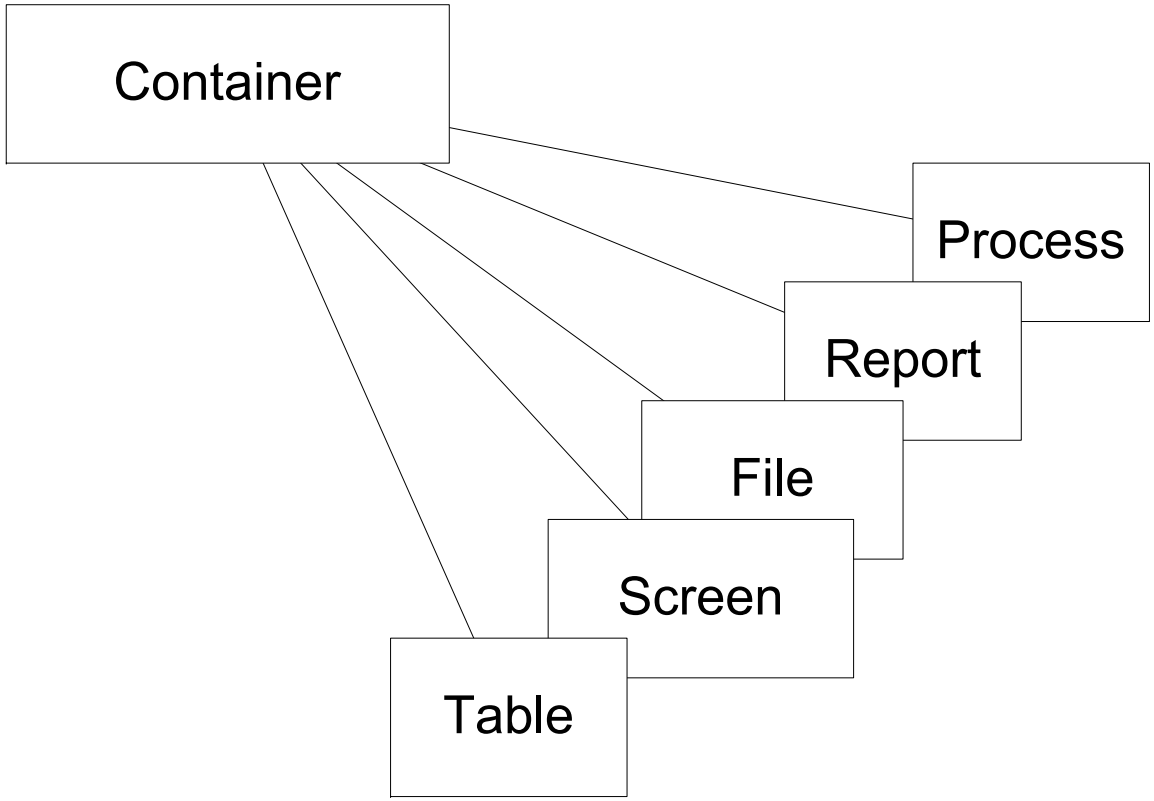
7.7 Business Domain (Not Prime Word!)

We need the business domain of the business fact, not it's container's name.

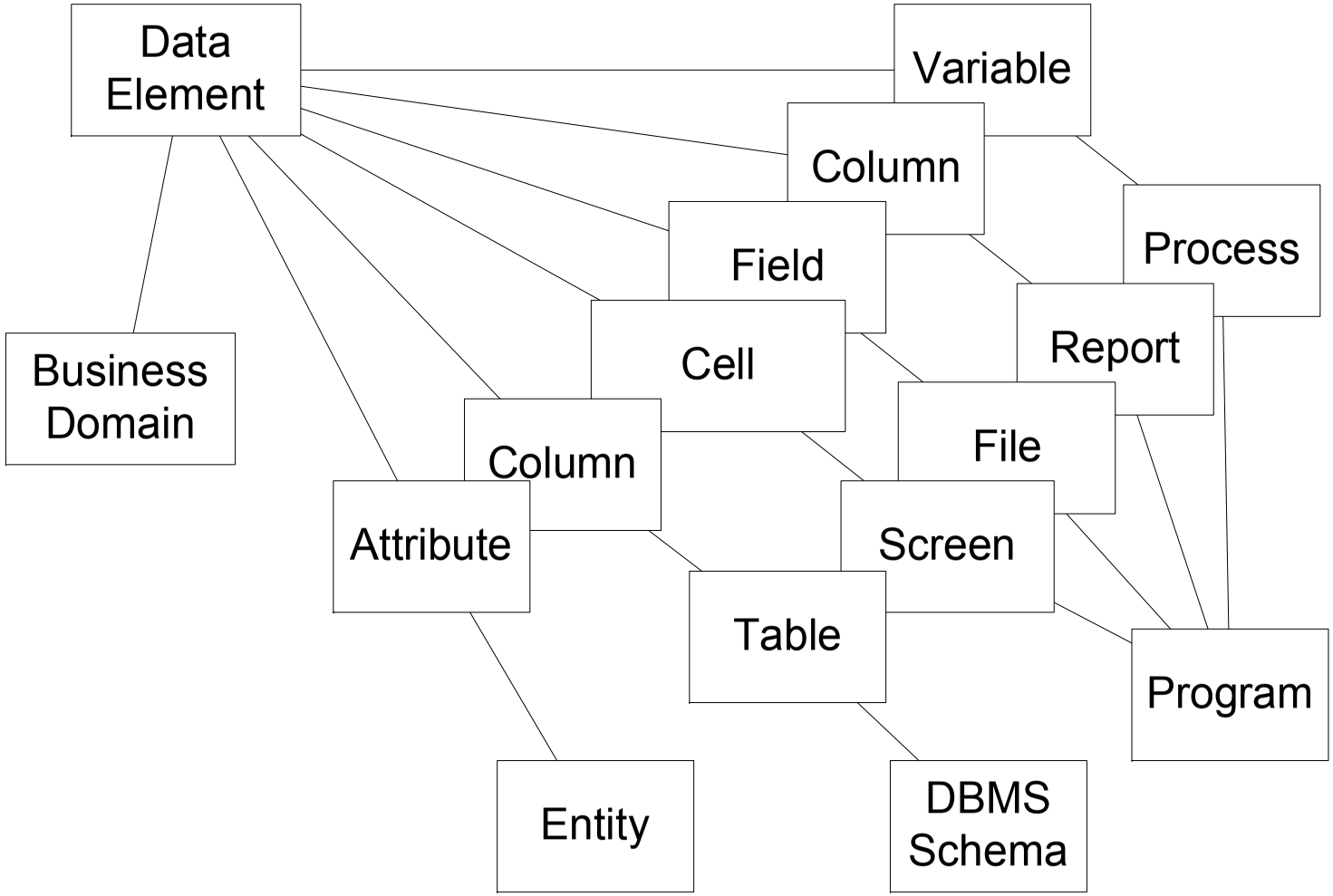
- Law–"contract vehicle id" is the contract number or the docket number
- Automobile–"motor vehicle id" is the vehicle identification number (VIN)
- Materials–"vehicle id" may be the identifier of the "medium"

In all three, the name may be the same "vehicle_id" but the business domain is essential to tell them apart.

Containers: Many classes of "containers" for incorporating the use of business facts...

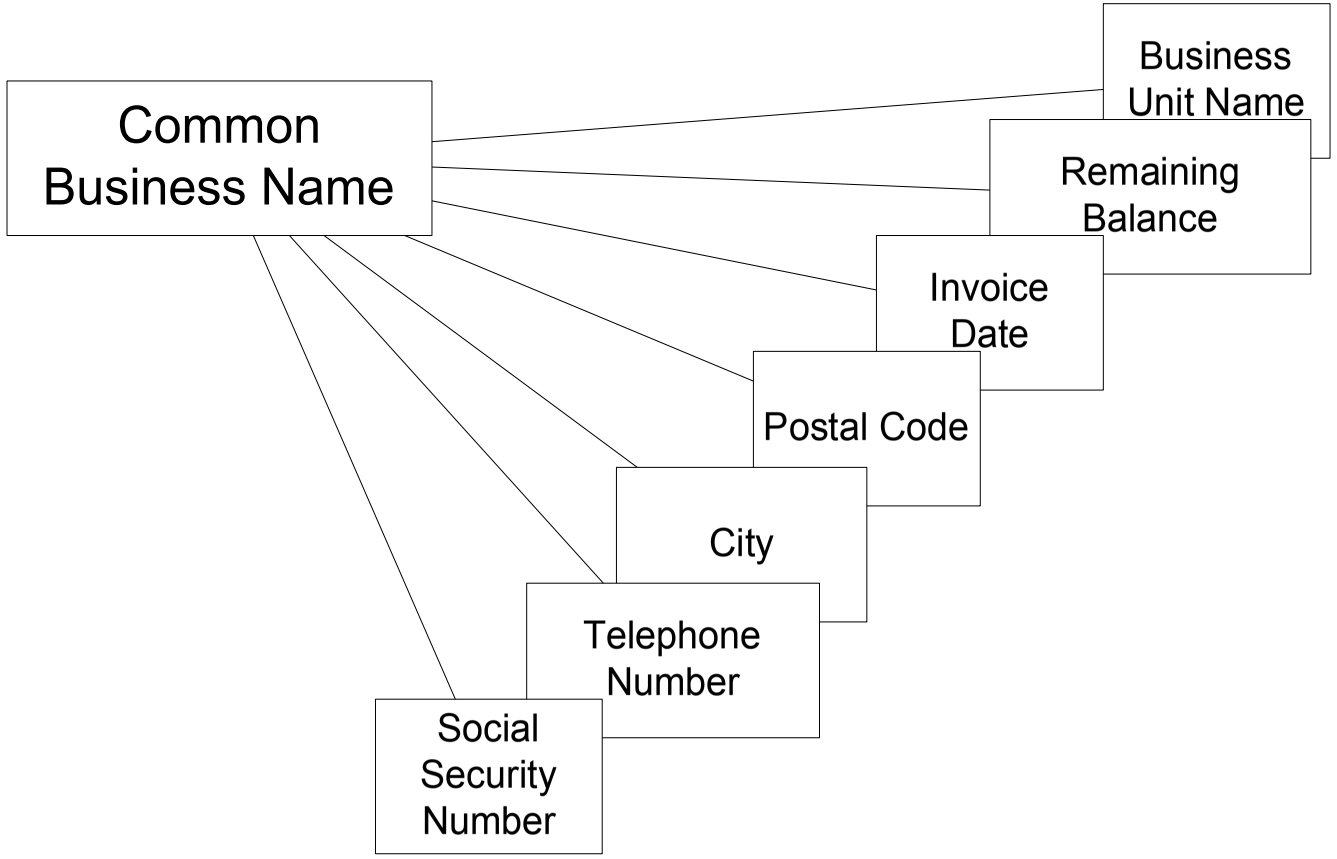


Essential Semantics of Business Facts are the same regardless of where and why they're used.



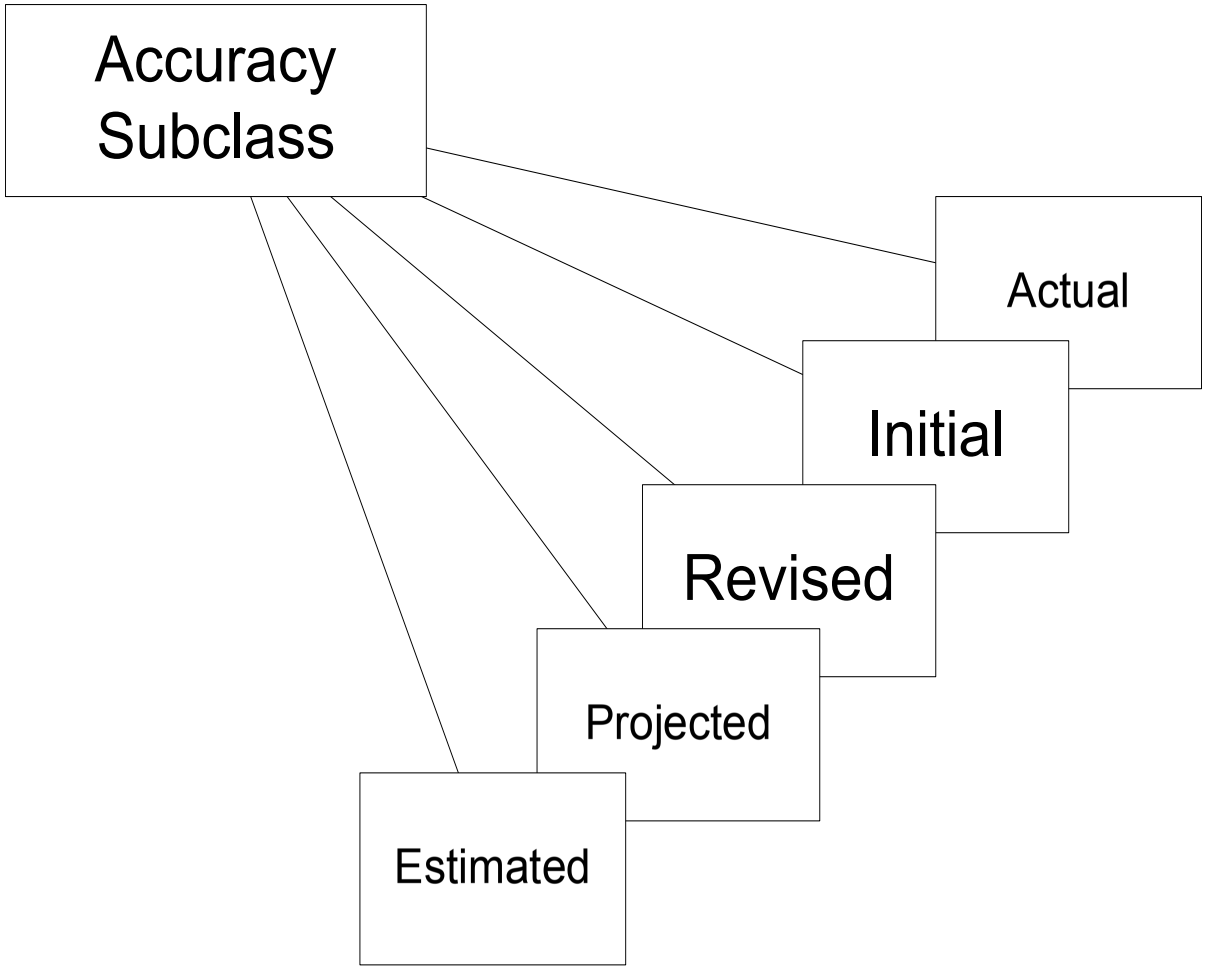
7.8 Common Business Name

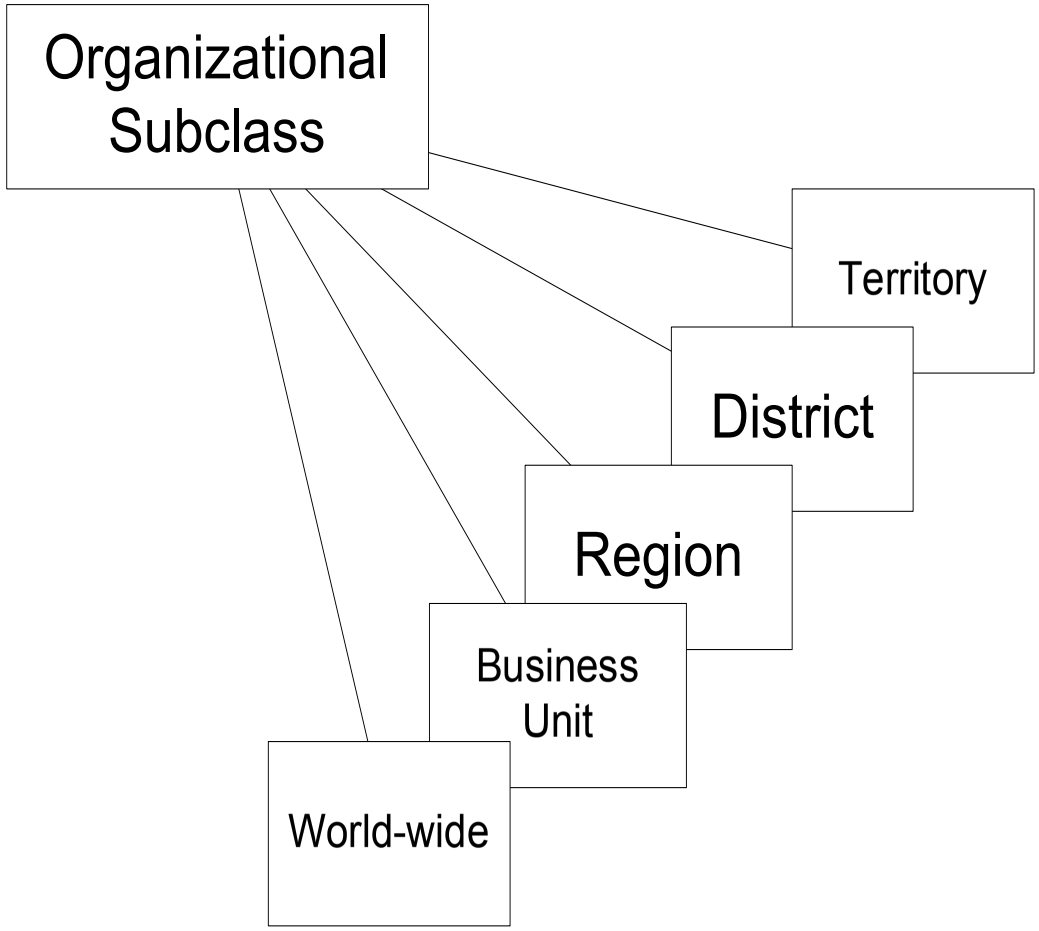
Every Business fact has a commonly used name. Don't change it to some arcane computer-based gibberish. USE THE COMMON BUSINESS NAME!

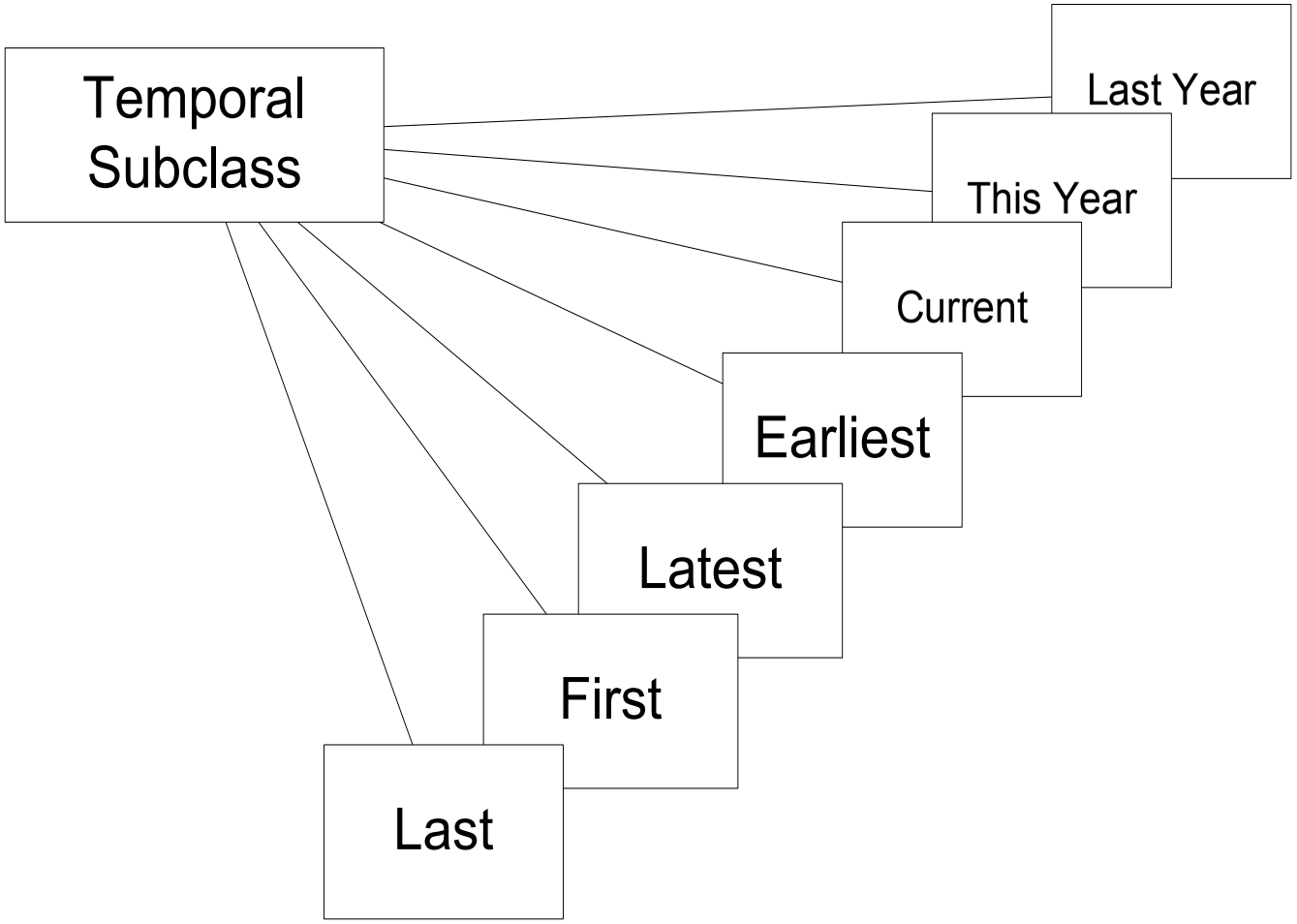


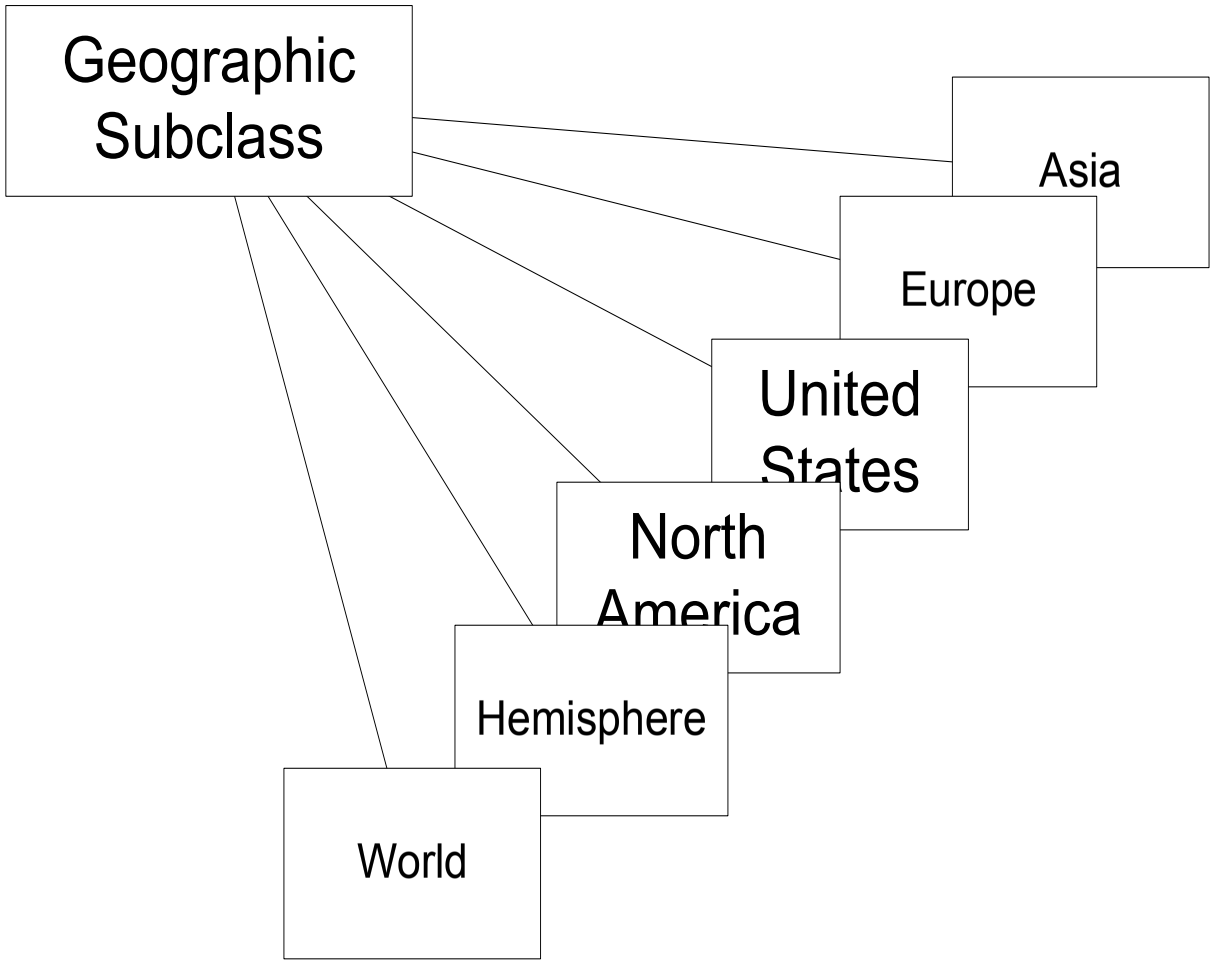
7.9 Modifier Classes

Modifier Classes and suggested contained values			
Temporal	**Accuracy**	**Geographic**	**Organizational**
last	estimated	world	world-wide
first	projected	hemisphere	business unit
latest	revised	North American	region
earliest	initial	United States	district
current	actual	mid-Atlantic	territory
this year		Maryland	
last year		Bowie	









For example,

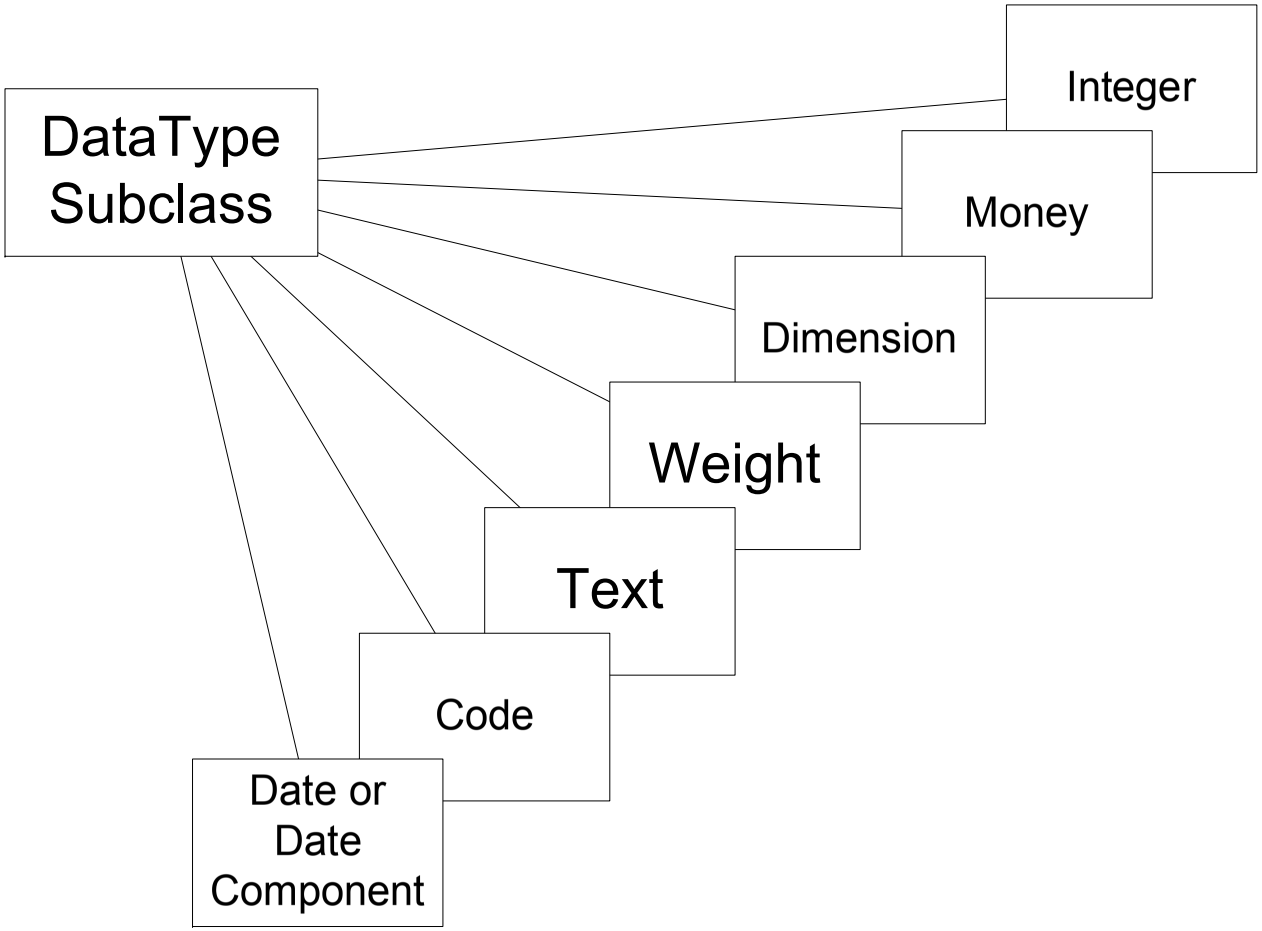
(world (hemisphere (North America (unites states (mid-Atlantic (Maryland (Bowie)))))))).

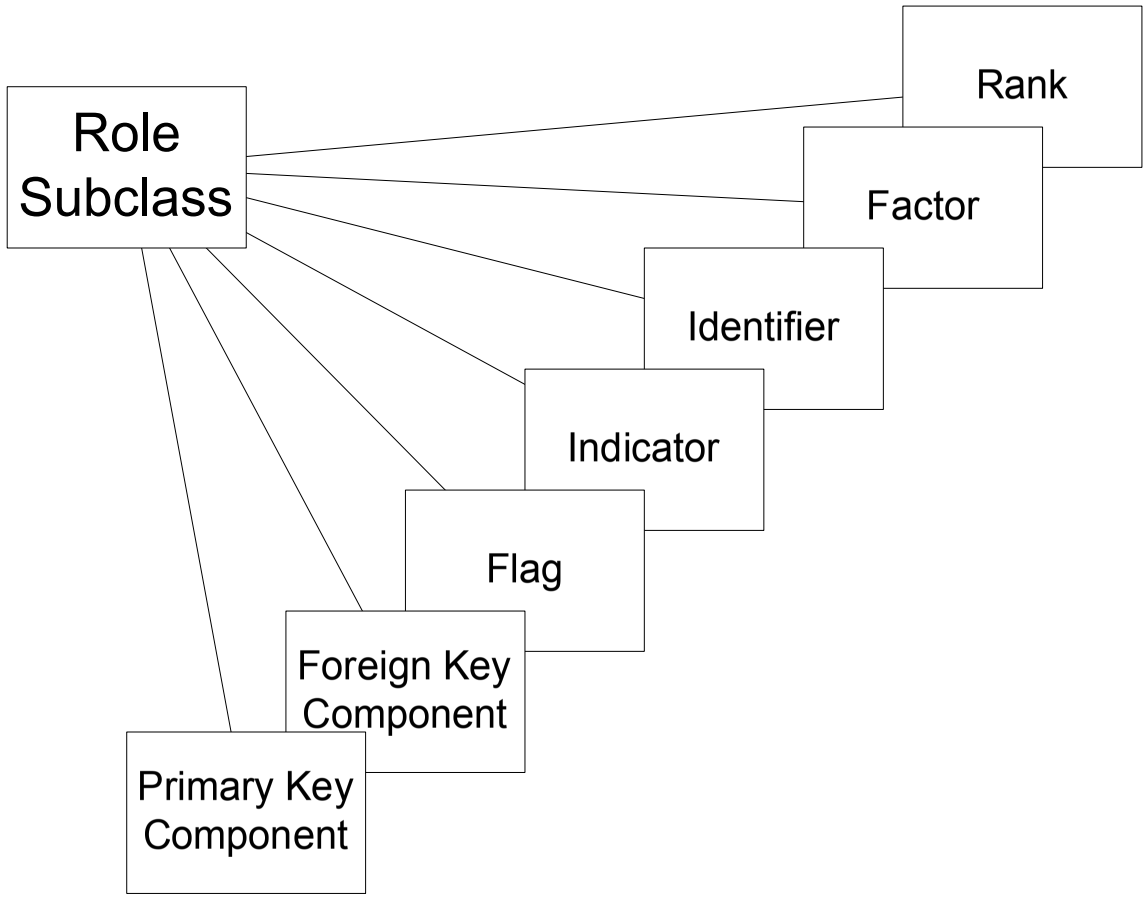
Regardless of how those arguments are made, the following both represent standard data and are thus semantically equivalent:

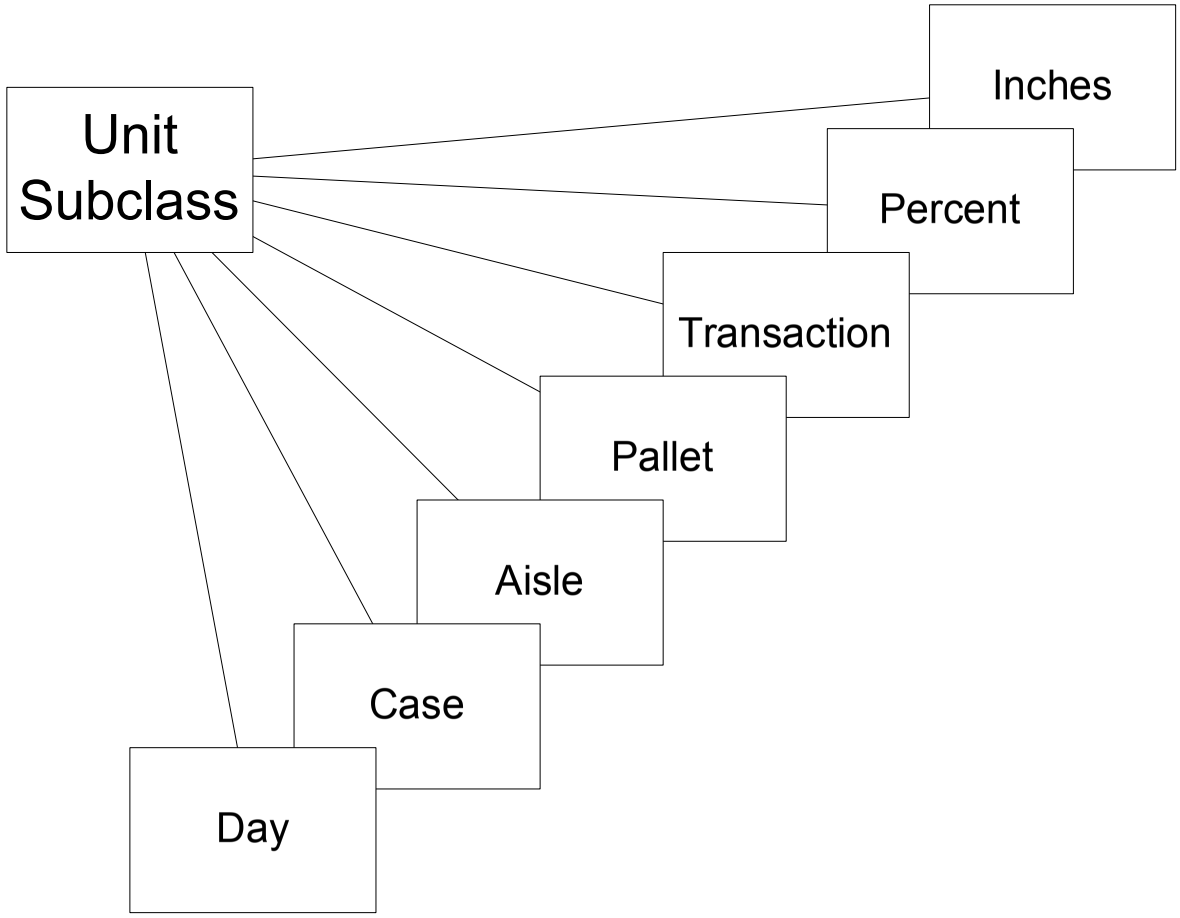
Columns	**Column values**
temporal	first
accuracy	projected
geographic	mid-Atlantic
organizational	regional
sales	$417,000,000

First_Projected_Mid_Atlantic_Regional_Sales	$417,000,000

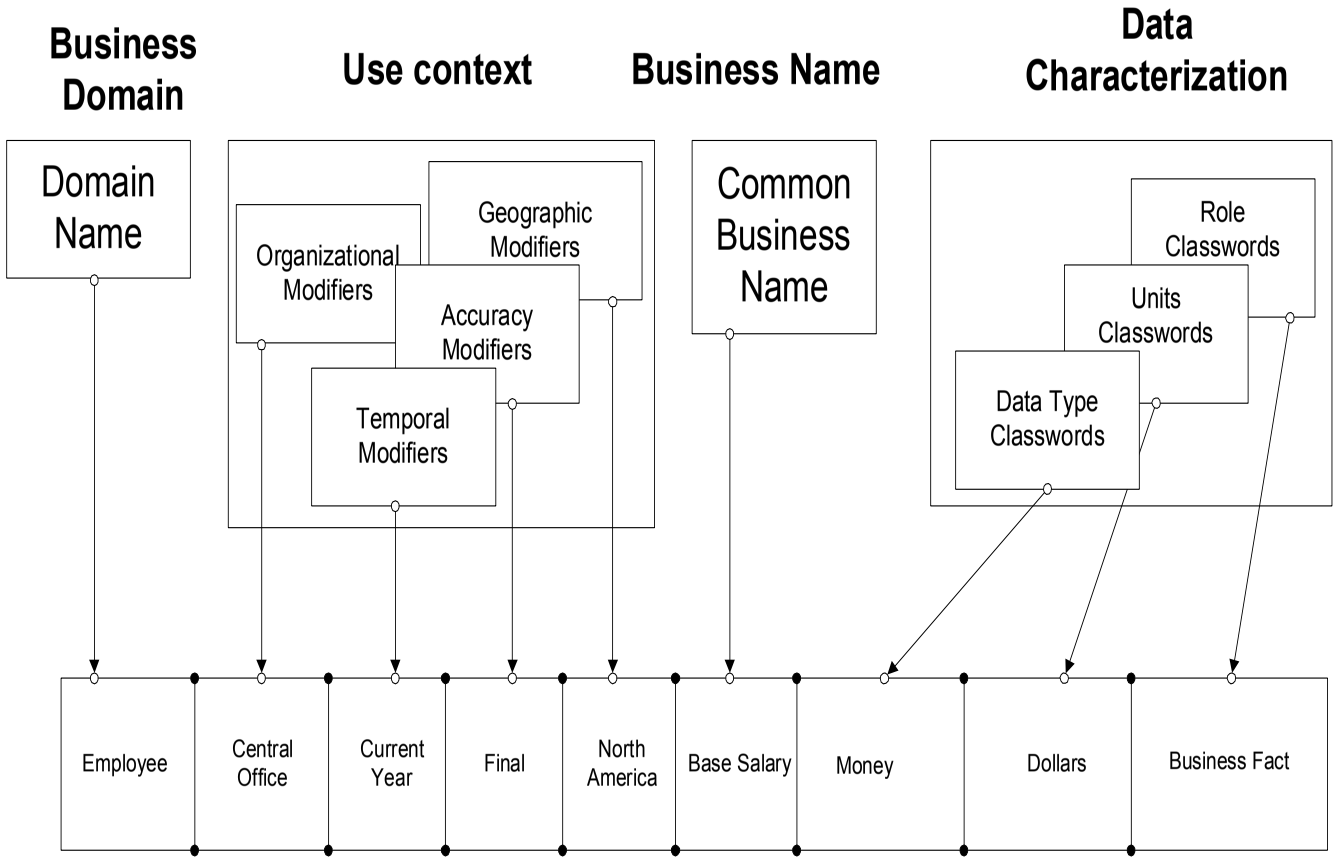
7.10 Class Word Classes







7.11 Full Data Element Name Construction



7.12 Semantics Summary

For Business Data Elements, the following is now known:

- Business domain
- Common business name
- Modifier set that define the business fact's geographic region, organizational unit, accuracy, and temporal characteristics
- Class word set for its role, units and basic data type

8.0 Value Sets

- Data values do not exist at random.
- Values fit patters
- Patterns are based on established policy
- There are restrictions on the values

8.1 Establishing Value Sets

- Determine the required maximum quantity of distinct occurrences
- Establish values that are not overlapping in meaning.
  For example, [O]verhand transportation, [R]ail, [T]rain.
- Establish values that do not contain conflicting semantics. For example, Employee status of [E]mployed, [A]ctive, [P]art time.
- Determine how the value is to be represented. That is, as a code (M or S), and/or as a full value (married or single), and possibly in multiple languages for example, Casado and Soltero.
- Determine if the value is simple or compound. If compound then divide the data field into its constituent data field parts.
- Determine if the value is also complex. If complex then divide the value into its simply contained parts including the definition of additional data fields to fully illuminate its meaning.
- Determine any data value characteristics such as maximums, minimums, allowed to be negative, and null.
- Estimate the maximum quantity of values that may exist in the foreseeable future

8.2 Objectives when Building Value Sets

- Consists of only simple values
- Can be readily expanded to meet foreseeable value growth
- The value set contains no obvious traps, such as only a two digit year code
- Determine if the values are to exist in pairs such as marital status name and marital status code. If so, then establish value pairs.

8.3 Value Set Maintenance

The maintenance procedure change document that contains:

- The values and meanings of the existing set of values
- The business policy problem that exists because of the existing value set
- The proposed modification of the existing value set
- The policy implication of adding, deleting, or modifying the existing value set
- The information systems technology impact on:
  - ♦ Existing data processing systems
  - ♦ Existing data
  - ♦ Historical data
  - ♦ Supporting documentation and training
  - A proposed time line for modification supported by resource estimates.

9.0 Data Standardization Cases

- **Simple**: single data element and restricted value set
- **Compound**: single data element but with multiply contained data elements
- **Group**: Single concept with multiple contained data elements such as address
- **Pair**: Two data elements that exist as matched pairs of codes and values
- **Related**: Two or more data elements that require one or more business rules to ensure consistency
- **Complex**: A data element that contain more than one of the preceding five data element situations

10.0 Data Standardization Work Breakdown Structure

Phase I. Data Standardization Project Planning

1 Perform overall data standardization project planning
1.1 Determine and achieve consensus on overall goals and objectives
1.2 Determine and achieve consensus on success measures
1.3 Determine and achieve consensus on evaluation measures
1.4 Review, revise, and achieve consensus on work breakdown structure (WBS)

1.5 Identify business unit involvement in project and phases

1.6 Identify and assign staff and accomplish project estimates
1.7 Acquire computing environment and training

2 Review, revise, and achieve consensus on all deliverable content and format for all comparison, difference, and resolution reports including
2.1 Identified data elements
2.2 Business domains of data elements
2.3 Existing data element characteristics such as business domain, and other semantics

2.4 Enterprise standard (international, regional, national, and local) data element characteristics
2.5 Difference between existing and enterprise standard

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

Phase II. Data Standardization Data Element Identification and Assessment

4 Identify, assign staff and accomplish phase estimate

5 Identify or develop mission and perform analysis
5.1 Identify or create overall mission for data standardization area
5.2 Create appropriate subordinate missions relevant to the data standardization area
5.3 Create mission for data standardization area
5.4 Create, store and validate through reporting the mapping between enterprise standard missions and the missions of the data standardization area
5.5 Create mission comparison, differences, and resolution report
5.6 Analyze report, rank issues, and make assignments for differences resolution
5.7 Identify relevant database domains

6 Identify or develop database domains and perform analysis
6.1 Select or create appropriate database domains
6.2 Create appropriate subordinate database domains relevant to the area of the data standardization area
6.3 Create database domains for data standardization area
6.4 Create, store and validate through reporting the mapping between enterprise

standard database domains and the database domains of the data standardization area

6.5 Create database domain comparison, differences, and resolution report

6.6 Analyze report, rank issues, and make assignments for differences resolution

6.7 Identify relevant database objects

7 Identify or develop database objects and perform analysis

7.1 Select or create appropriate database objects

7.2 Analyze the scope, purpose and coverage of the data standardization area's database objects

7.3 Create, store and validate through reporting the mapping between enterprise standard database objects and the database objects of the data standardization area

7.4 Create database object comparison, differences, and resolution report

7.5 Analyze report, rank issues, and make assignments for differences resolution

8 Identify or develop database object data structure analysis

8.1 Select or create appropriate database object data structures

8.2 Analyze the scope, purpose and coverage of the data standardization area's database object data structures as evidenced through its tables and/or files

8.3 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

8.4 Create database object data structure comparison, differences and resolution report

8.5 Analyze report, rank issues, and make assignments for differences resolution

8.6 Perform standard data element analysis

8.7 Select or create appropriate data elements

8.8 Select or create semantics for data elements

8.9 Identify data standardization area's data element deployment

8.10 Create, store and validate through reporting the mapping between standard data elements and the deployed uses of the data elements within the data standardization area

8.11 Create a comparison, difference, and resolution report for international, regional, national, and local data

8.12 Analyze report, rank issues, and make assignments for differences resolution

9 Perform deployed data element analysis

9.1 Select or create appropriate deployed data elements including

9.2 Select or create semantics for deployed data elements

9.3 Identify data standardization area's deployed data elements

9.4 Create, store and validate through reporting the mapping between standard data

element and the deployed data elements within the data standardization area

9.5 Create a comparison, difference, and resolution report for international, regional, national, and local deployed data elements

9.6 Analyze report, rank issues, and make assignments for differences resolutionPerform business policy research and formulation

9.7 Identify current policy basis for the deployed data element

10 Review and/or formulate data definitions and standards for deployed data elements

10.1 Identify all deployed data element that are intended to embrace the same set of semantics and values by business unit, computing system, and database

10.2 Analyze the semantics within the deployed data element set to determine differences

10.3 Analyze the values within sets of deployed data elements to determine differences

10.4 Determine the mapping between semantic differences for deployed data elements

10.5 Determine the mapping between value set differences deployed data elements

Phase III. Data Standardization Implementation

11 Identify, assign staff and accomplish phase estimate
11.1 Identify phase project manager
11.2 Determine members of phase team
11.3 Develop detailed phase estimate
11.4 Accomplish resource loading, build PERT, Gantt and CPM charts for phase
11.5 Present phase plans and revise as necessary
11.6 Identify and assign administrative support
11.7 Identify and acquire automation/tools support

12 Accomplish Data element Standardization
12.1 Obtain the report from data element standardization
12.2 Determine the resources (hardware, software, peopleware, and time) for required policy changes
12.3 Determine the resources required for operating system changes
12.4 Determine the resources required for existing database changes
12.5 Determine the resources required for historical data system changes

12.6 Determine the intra and inter business unit risk for not accomplishing data standardization

13 Assignment of management for review and approval
13.1 Present the report that identifies, analyzes, and costs the effect of nonstandarization of critical data elements
13.2 Present the report that identifies the costs of a critical data element data standardization effort
13.3 Present the recommendation for data standardization
13.4 Obtain a management decision to remain the same or to proceed with data standardization effort
14 Proceed with data standardization implementation project
14.1 Plan project
14.2 Accomplish the database changes
14.3 Accomplish the existing computing system changes
14.4 Accomplish the policy and procedure changes
14.5 Accomplish the actual data value migration for current and historical data
14.6 Deliver the project results and develop lessons learned

11.0 Summary

Because of this approach to data standardization,

- Enterprises will be able to properly focus on the data concepts such that the lowest level column is easily and automatically named.

- Enterprises will have data contexts are immediately apparent without compromising data identification, selection, and melding.

- Enterprises will be able to accommodate multiple implementation technologies such that regardless of the rules, a data concept are immediately obvious and relatable to all other data that espouse the same concept.

- Enterprises will be able to empower their front-line project staff to create and maintain their own names under the guidance and work enhancing tools and techniques of a central standardization and maintenance authority