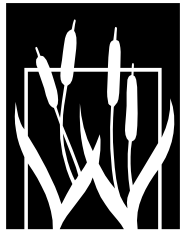


Whitemarsh
Information Systems Corporation

A Column by Any Other Name Is Not A Data Element

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

Objective of Talk 1
Fundamental Data Element Definition 1
Test of this Approach 1
A Michael Brackett Story from DAMA 2001 2

The Approach 3

ISO 11179, The Data Element Metadata Standard 4
Administration and Identification 5
Conceptual [Value] Domains 7
Data Elements 11

Melded ISO 11179 Compliant Data Element Meta Model 12

Benefits to this Approach 13

Objective of Talk

- Describe an approach to achieve enterprise-wide data standardization
- Through the specification, implementation, and maintenance of data elements
- Within the context of a metadata-repository, CASE-like environment

Fundamental Data Element Definition

- Data elements are context independent business fact semantic template
- That are employed to fully define and control context dependent business facts
- Such as attributes of entities, columns of tables, fields on forms, etc.

Test of this Approach

- "Does this approach make common sense?"
- Is the demonstration compelling?
- If yes, then consult references for the much deeper presentations

A Michael Brackett Story from DAMA 2001

- Michael Brackett sat in on the my Enterprise Wide Data Standardization presentation at DAMA 2001 in Anaheim.

- I "creatively acquired" significant components of the Whitemarsh approach from a May 1995 Michael Brackett talk

- It was sort of like having Michelangelo sit in our your one-person art show.

- You're both wanting his review and critique but are deathly afraid that he will give it, or even worse, just walk out half way through. I got the critique and review, and Michael stayed to the very end.

- After the presentation was over, a meeting with Michael produced the observation that the approach was fundamentally sound but that he disagreed with my assertion that an enterprise only had 2000 data elements. Wow, I thought, he agrees with the approach. That's great!

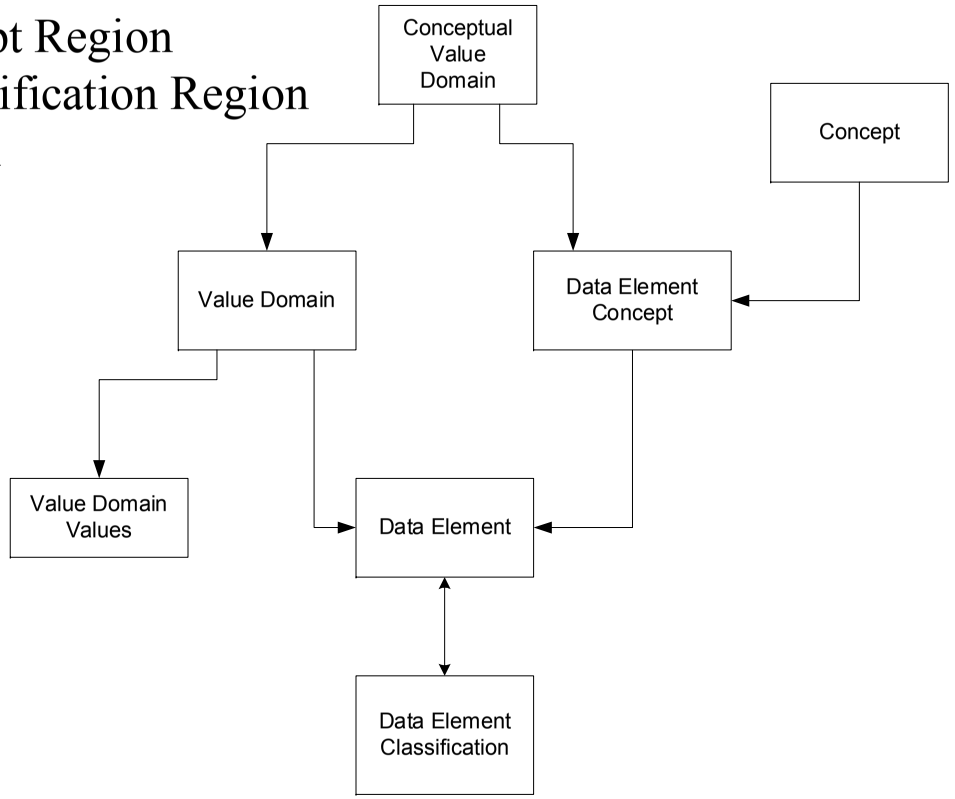
- Brackett stated that the state of Washington, for example, has about 20,000 data elements. A casual review of a "business" like the State of Washington is between 10-15 different enterprises (Agriculture, Education, Environment, Justice, Transportation, Welfare, etc.)

The Approach

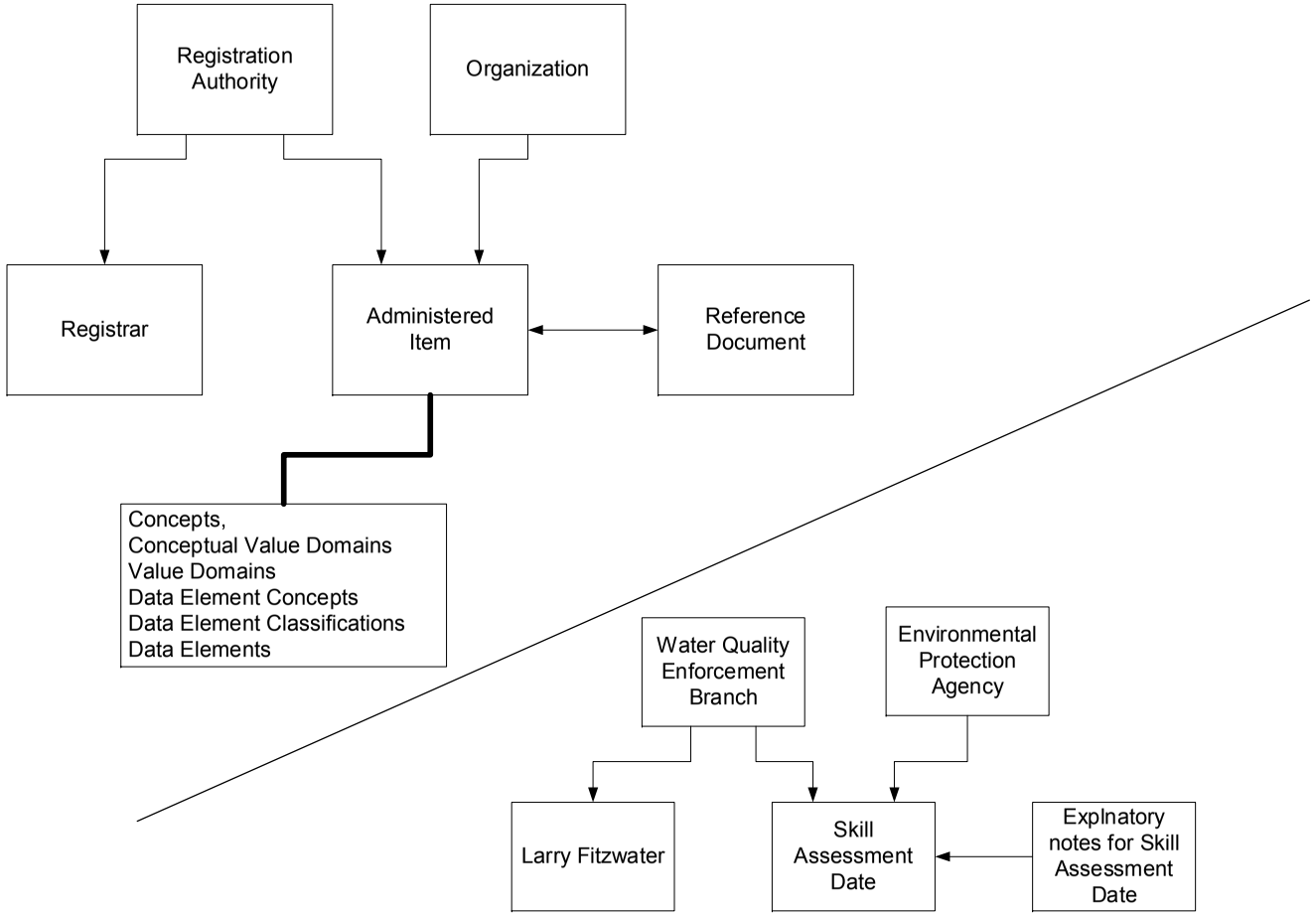
- Create a re-usable cache of data elements
- Establish a metadata repository and CASE environment
  - Define database columns,
  - Interrelate the defined columns through the data element metadata, and
  - Support both forward engineering of new databases and reverse engineering of existing databases.

ISO 11179, The Data Element Metadata Standard

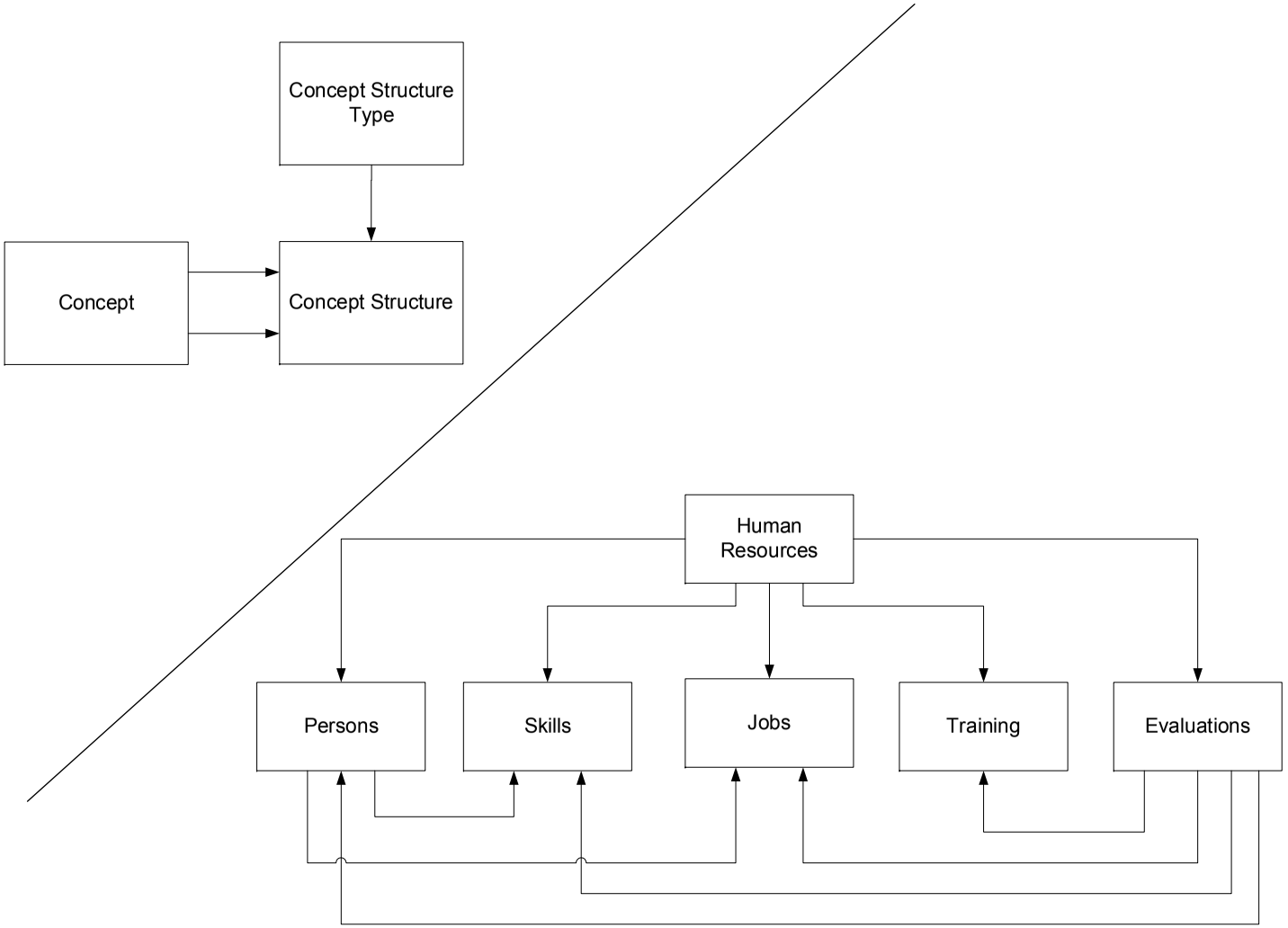
- Administration and Identification Region
- Concepts
- Conceptual [Value] Domain
- Value Domains
- Data Element Concept Region
- [Data Element] Classification Region
- Data Element Region



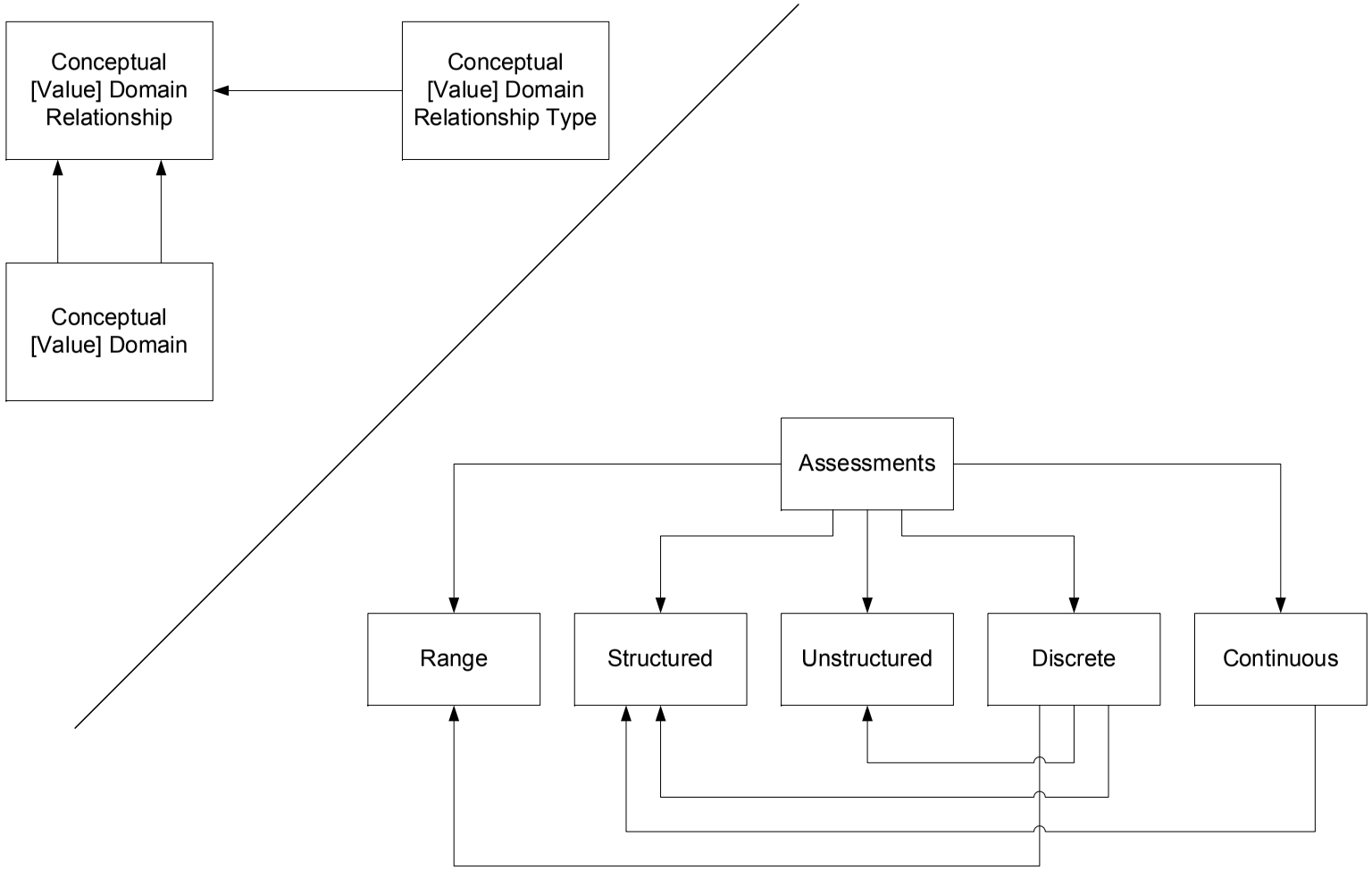
Administration and Identification



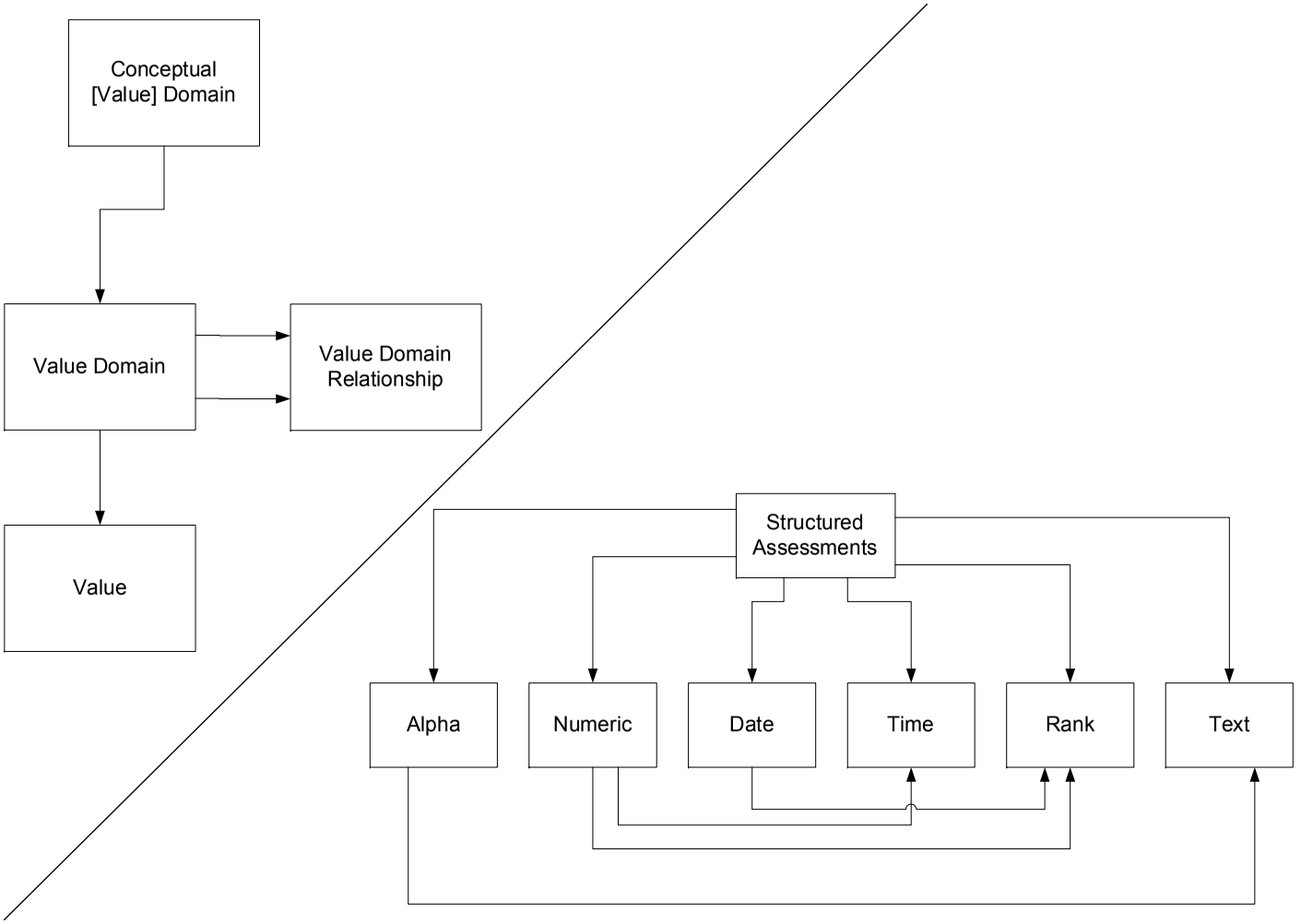
Concepts



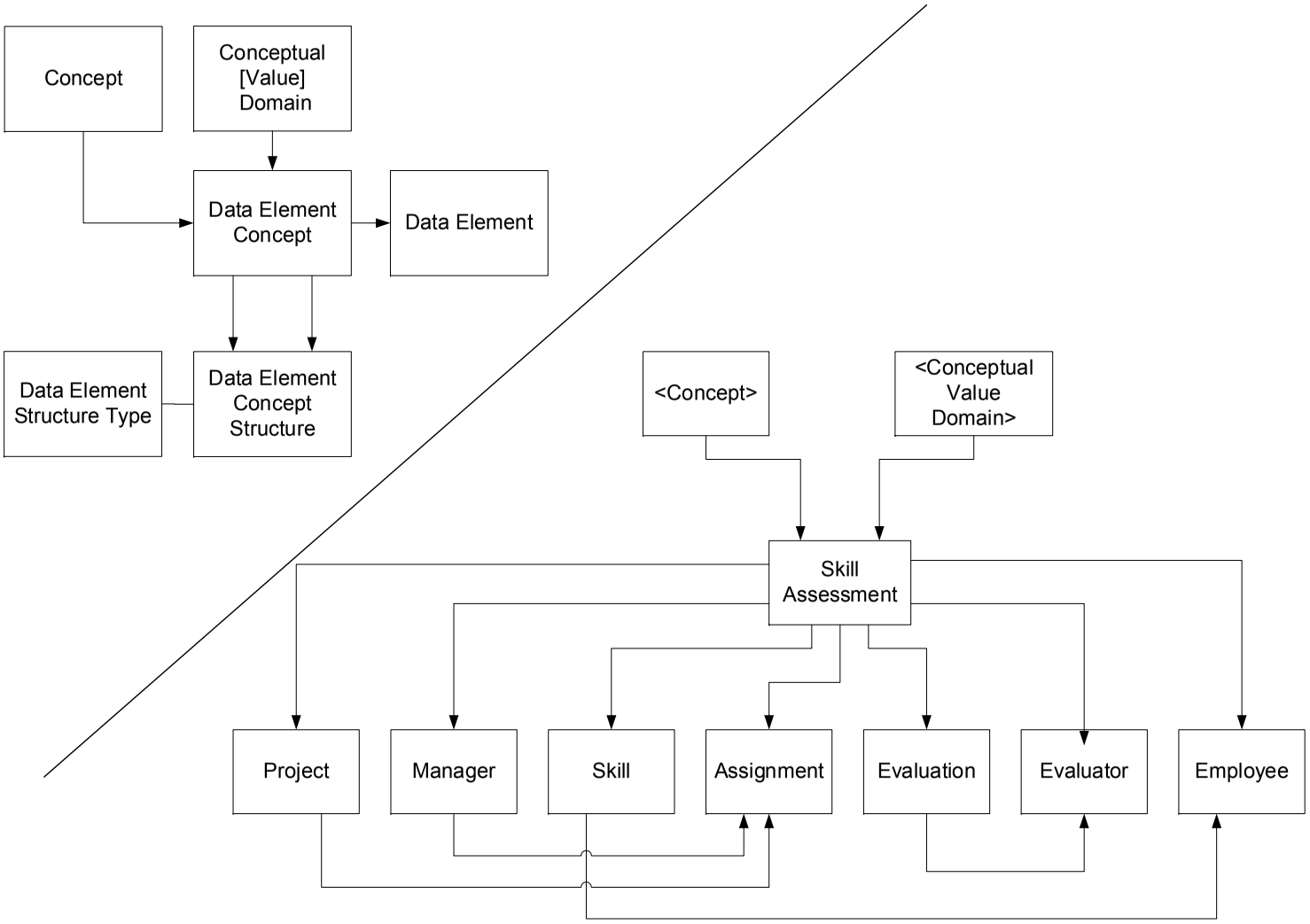
Conceptual [Value] Domains



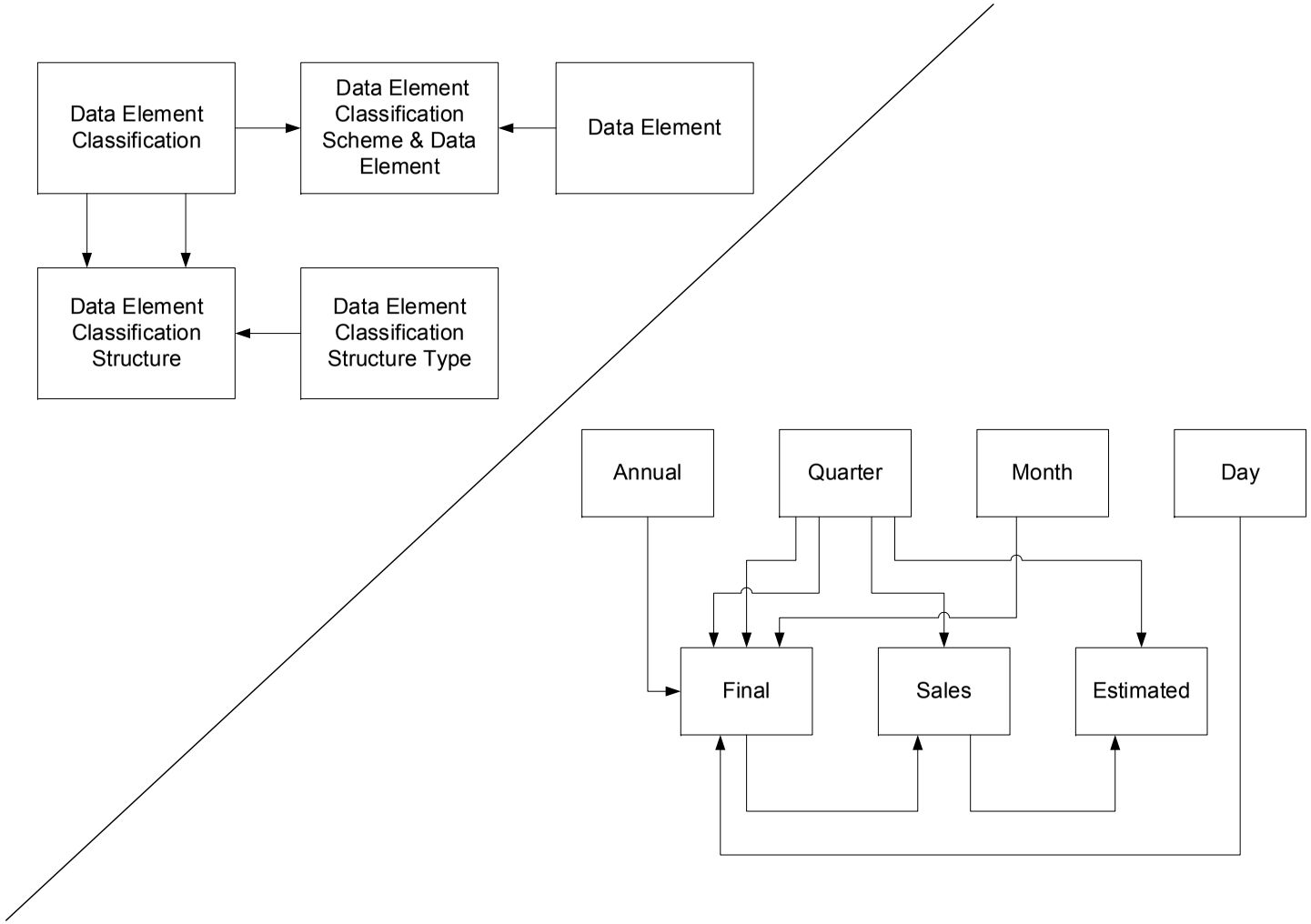
Value Domains



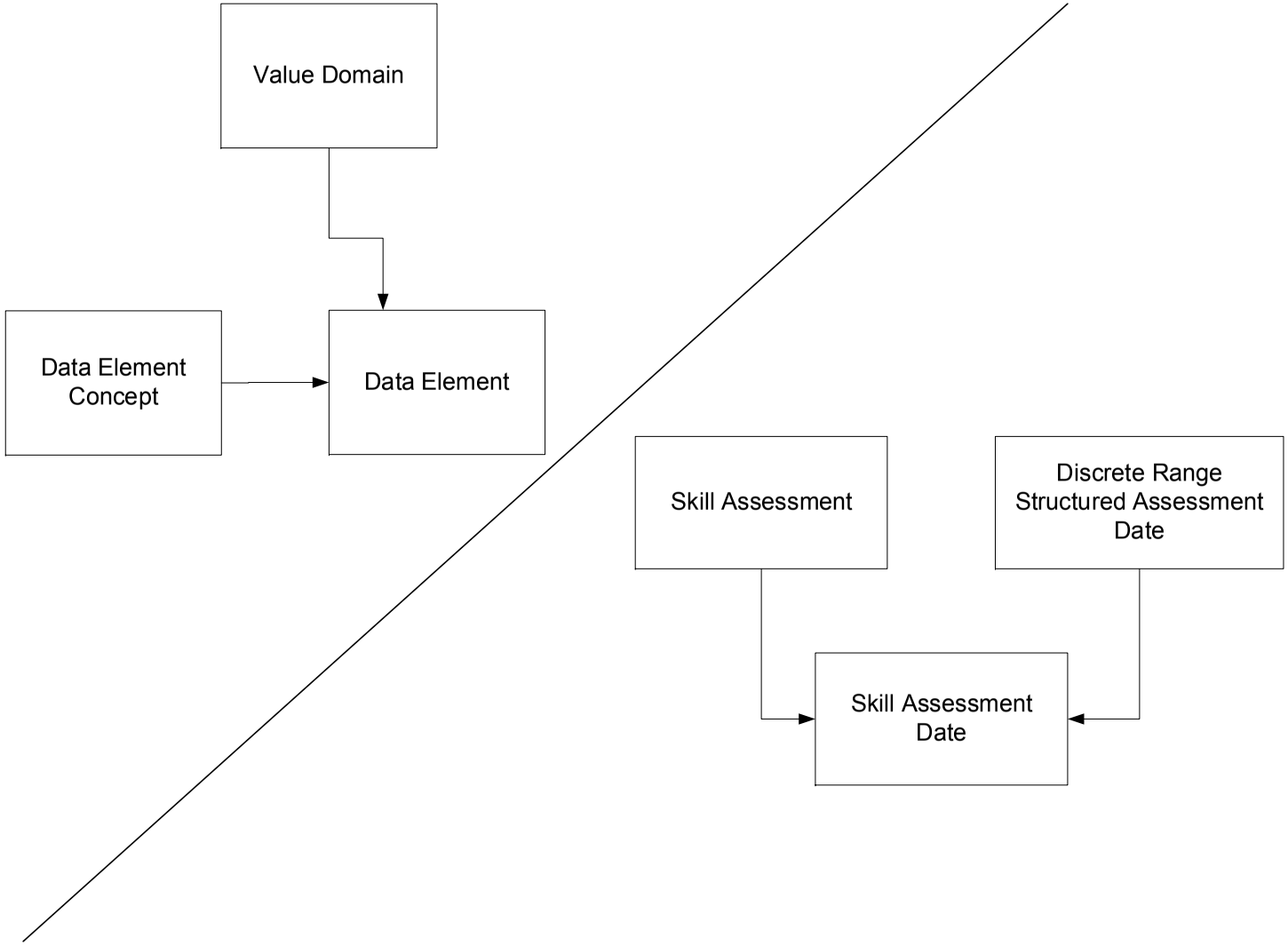
Data Element Concepts



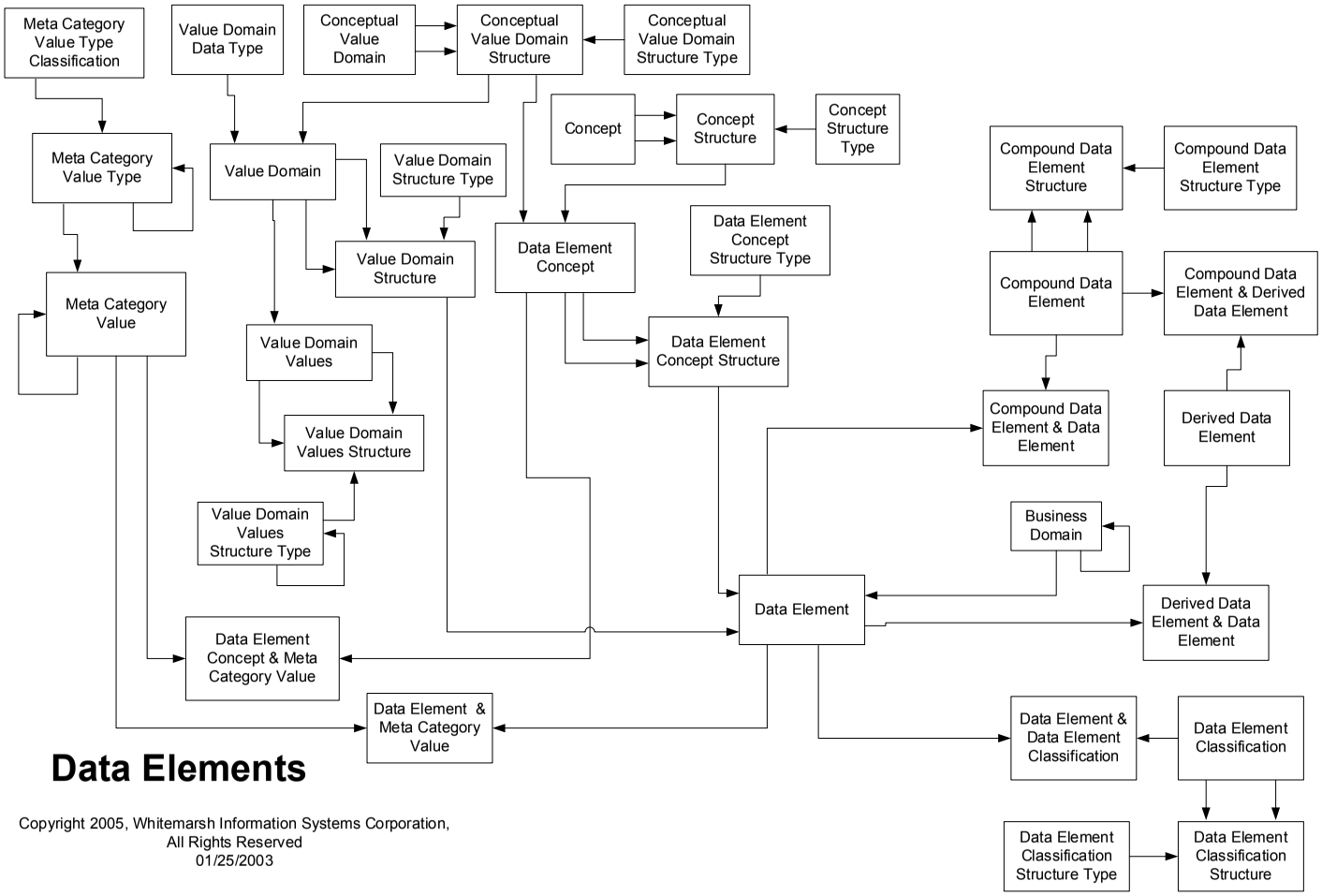
[Data Element] Classifications



Data Elements



Melded ISO 11179 Compliant Data Element Meta Model



Benefits to this Approach

Common statistics about database environments for a typical state government, or multiple business line enterprise.

Unit	Components
1 database	100 tables
1 table	15 columns
Each agency	100 databases
Each state	40 agencies

- The total count of columns for each database is 1500.
- The total columns for an agency is 150,000 columns,
- The total for the state, for example, like Washington, New York, or Maryland would be 6,000,000.

Thus, if Michael Brackett's assertion is true, then each of the asserted 20,000 data elements is reused about 300 times.

Alternative Approaches and Cost Comparisons		
Data Standardization Alternative for a multi-site, multi-application Government MIS	**Final Quantity of columns, fields, cells, etc.**	**Cost via technique employed for definition**
Accomplished traditionally (prime + modifier + classword) across all systems	19,000	$6.75 million
Alternatively, if accomplished by standardizing closely named columns and fields	3,000	$1.06 million
Alternatively, if accomplished through Comprehensive Data Standardization Techniques–Eliminates redundant–but different-- representations of the same concept	560	$200,000

In this example the ratio was 34 to 1.
Another Example: US Department of Defense Agency ETL Effort

- Each: requirement, design, software implementation and maintenance.
- Each ETL represents a failure in data standardization.
- Columns supposed to be the same have different names, semantics, data types, levels of granularity, time-sequencing, and the like.
- While an enterprise-wide data element standardization approach would not solve all these problems, it would clearly affect different names, semantics, data types.
- The agency spends about $175,000,000 each and every year on such ETL activities.
- If the data element approach resolved 50% then that would represent savings of about $90 million per year.
- Extended to the US DoD as a whole the savings would be about $450 million, and to the U.S. Government as a whole, about $1.5 Billion.

- Given that the US Government spending represents about 10% of the total economy, then the savings to the economy as a whole is about $15 Billion.