



Whitemarsh  
Information Systems Corporation

## *Data Management Program Way Ahead*

### *A Case Study: United States Army*

Whitemarsh Information Systems Corporation  
2008 Althea Lane  
Bowie, Maryland 20716  
Tele: 301-249-1142  
Email: [Whitemarsh@wiscorp.com](mailto:Whitemarsh@wiscorp.com)  
Web: [www.wiscorp.com](http://www.wiscorp.com)

## Table of Contents

|   |     |
|---|-----|
| Acknowledgments .....   | iii |
| Data Management Program Mission .....   | 1   |
| 1.0 Current Data Standardization Alternatives .....   | 2   |
| 1.1 Interfaces .....  | 2   |
| 1.2 DDDS (DoD Dictionary System) .....  | 4   |
| 1.3 The XML Alternative to “Data Standardization” .....   | 6   |
| 2.0 Alternatives Common Characteristics .....   | 8   |
| 3.1 Build Five Levels of Abstraction to Enable Inter-community-of-interest Harmonization .....                | 11  |
| 3.2 Mine the DDDS and DDA for Existing Consensus on Shared Data Elements and Shared Data Segments .....       | 12  |
| 3.3 Perform Data Standardization within Communities of Interest .....   | 13  |
| 3.4 Standardize Only at the “Shared Intersections.” .....   | 14  |
| 3.5 As Data Standardization Advances, Data Models & Metadata Converges .....                                  | 15  |
| 3.6 Build infrastructure to “automate” database design, and generate XML Schemas and Discovery Metadata. .... | 16  |
| 3.7 Build Infrastructure to Consolidate or Eliminate Systems. ....  | 17  |
| 3.7.1 Consolidate Databases into Fewer and More Expansive .....   | 18  |
| 3.7.2 Re-cast Monolithic Systems into More Agile, Distributed Systems. ....                                   | 21  |
| 4.0 Applicability to Prototypical System .....  | 24  |
| 4.1 Cost of Data Model Development .....  | 24  |
| 4.2 Cost of Prototypical System Software .....  | 25  |



## **Acknowledgments**

This material is an evolution of a document accomplished during the time frame: September 2003 through December 2004. The primary authors were Bruce Haberkamp, James Blalock, and Michael Gorman of the Office of the CIO, United States Army. The foundational components of this work has been favorably reviewed by subject matter experts within the U.S. Department of Defense.



## Data Management Program Mission

*Enable*

*Understanding-Based Data Interoperability*

*throughout the Enterprise*



## 1.0 Current Data Standardization Alternatives

### 1.1 Interfaces

- The USAF 1995 Study on Interface costs showed that each interface costs about \$335,000 per year to create and maintain.
- The USAF was extrapolated to then be spending about \$167 million per year (1995 dollars).
- If the USAF is 20% of DoD, then DoD is spending about \$1 Billion per year on just IT interfaces.
- Given that the Army has 20% of the DoD's IT then it's cost is about \$200 Million per year for interface activities.
- All interoperability is on a point-to-point basis.



- We need to unbind interfaces from 1-off to generalized
- Scaled to make the entire DoD Interoperable, then
  - ◆ If every 30 systems have 100 interfaces, and Navy has 75,000 systems, then there are 250,000 interfaces in the Navy. That's about 1 million interfaces in the DoD.
  - ◆ On a per year basis, each interface is \$1/3 million, thus annual cost = \$335 Billion



## 1.2 DDDS (DoD Dictionary System)

An alternative to point-to-point interfaces is “8320 Data Standardization” That is, the DDDS (DoD Dictionary System) and the Defense Data Architecture (DDA). DDDS failed for many reasons. It could never have succeeded because of its engineering!

- The U.S. Navy has 75,000 systems.
- Assume the Navy has 50,000 databases across these 75,000 systems.
- If each database has 100 tables , and each table has 15 columns, then
- The Navy has  $50,000 * 100 * 15$  columns, or 75,000,000 database columns to standardize.
- If the Navy is 20% of DoD then DoD has 375 million columns (DDDS Data Elements) to standardize.



- At two staff hours per “data element” the cost to define and standardize would be \$56.25 Billion
- DDDS was halted at about 20,000 approved “data elements.” Thus, they had completed about 0.005% of the work.
- There still wouldn’t have been standardization because every column would have been its own standard.





### 1.3 The XML Alternative to “Data Standardization”

XML was started as the “antidote” to Data Standardization

- Suppose, again, the U.S. Navy has 75,000 systems.
- Each system has 100 screens. Each screen has 3 sub-screens.
- Each sub-screen’s data specification has to be represented by a XML Schema.
- Thus, the Navy has a need for  $75,000 * 100 * 3$  XML schemas, or 22,500,000 XML schemas.
- DoD as a whole then has a need for 112.5 million XML schemas.
- If each XML schema has 15 XML elements, then the needs to be 1.688 Billion XML element definitions.



- That's 4.7 times LARGER effort than the DDDS effort that DISA *stopped*.
- Further, suppose that each XML schema takes 8 staff hours to define.
- DoD agencies are required to then spend, 430,769 staff years defining their XML schemas.
- At \$150K per staff year, that's a total cost of about \$64.6 Billion.



## 2.0 Alternatives Common Characteristics

- None achieve their objective of creating understanding-based data interoperability because every “data element has a 1-off definition.”
- All are really point-to-point interfaces because they all “standardize” at the column level.
- All have unacceptable price tags
  - ◆ Point-to-point – \$335 Billion per year
  - ◆ DDDS – \$52.25 Billion (without any evolution, maintenance or configuration management)
  - ◆ XML – \$64.4 Billion (without any evolution, maintenance or configuration management)
- None have any “reuse” potential.



- All metadata is isolated and is “system centric.” No enterprise orientation.
- Finding the right XML Schema from among 112.5 million is like finding a snow-flake in a blizzard.
- No possible strategy for configuration management, maintenance/evolution, or where-used analyses.
- Once scaled, all three alternatives are cost prohibitive.

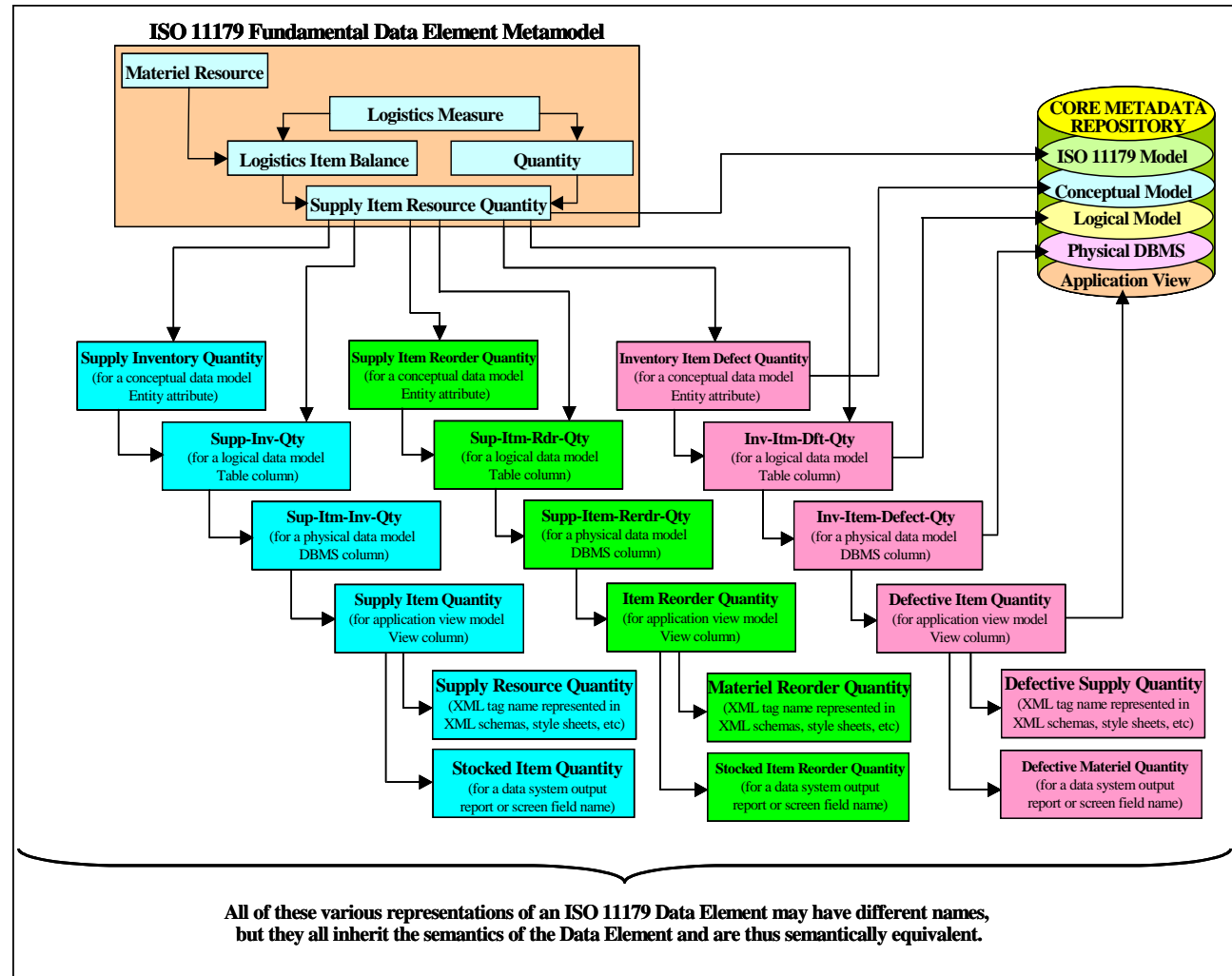


### **3.0 Data Management Program Way Ahead**

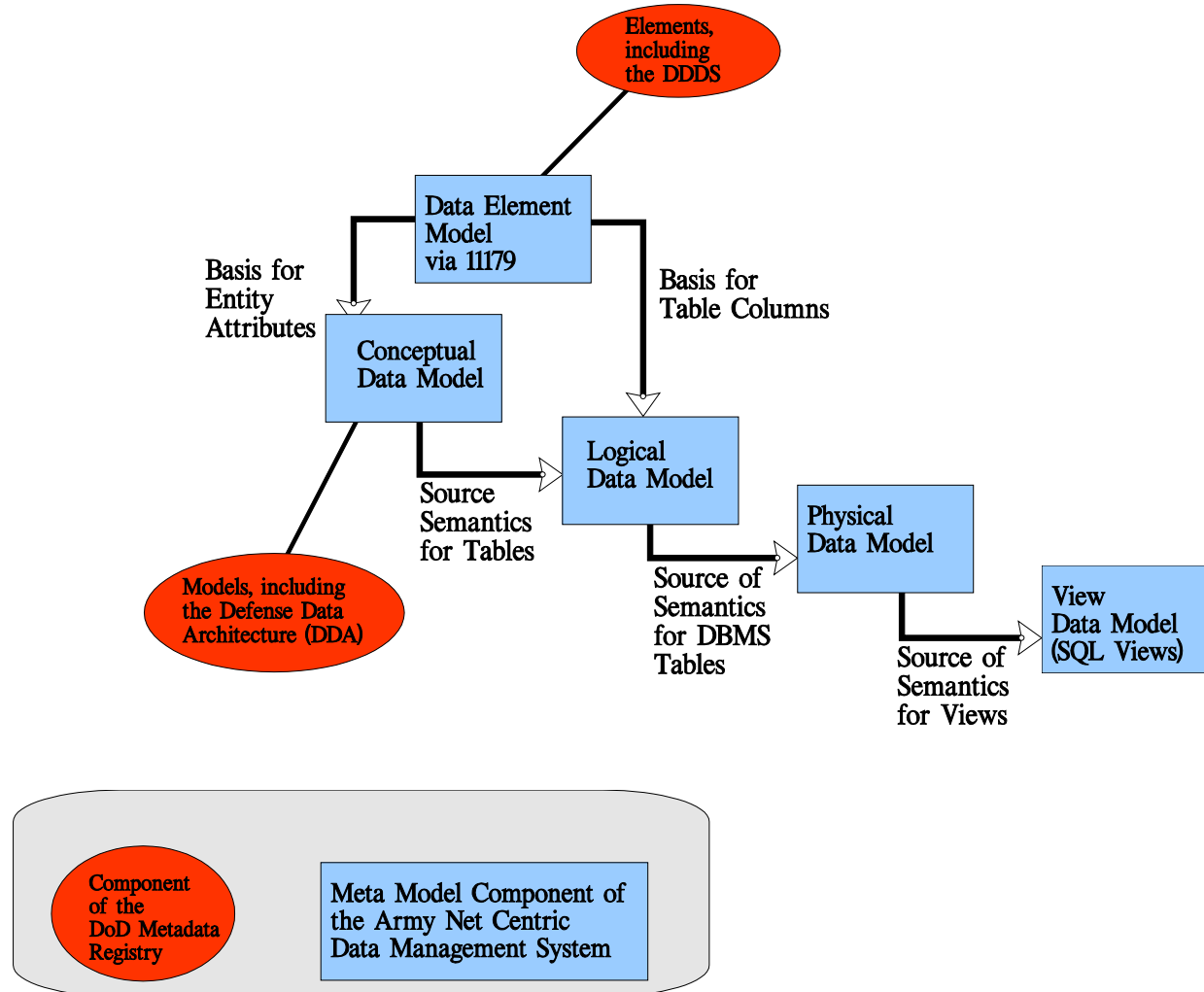
1. Build five levels of abstraction to enable Inter-Community-of-Interest Harmonization.
2. Mine the DDDS and DDA for existing consensus on Shared Data Elements and Shared Data Segments to quick-start Inter-COI-Harmonization.
3. Perform Data Standardization within Communities of Interest.
4. Standardize only at the “shared intersections.”
5. As Data Standardization Advances, Data Models and Metadata Converges
6. Build infrastructure to “automate” database design, XML Schema generation, and Discovery metadata.
7. Build infrastructure to consolidate or eliminate systems.



### 3.1 Build Five Levels of Abstraction to Enable Inter-community-of-interest Harmonization



### 3.2 Mine the DDDS and DDA for Existing Consensus on Shared Data Elements and Shared Data Segments



### 3.3 Perform Data Standardization within Communities of Interest

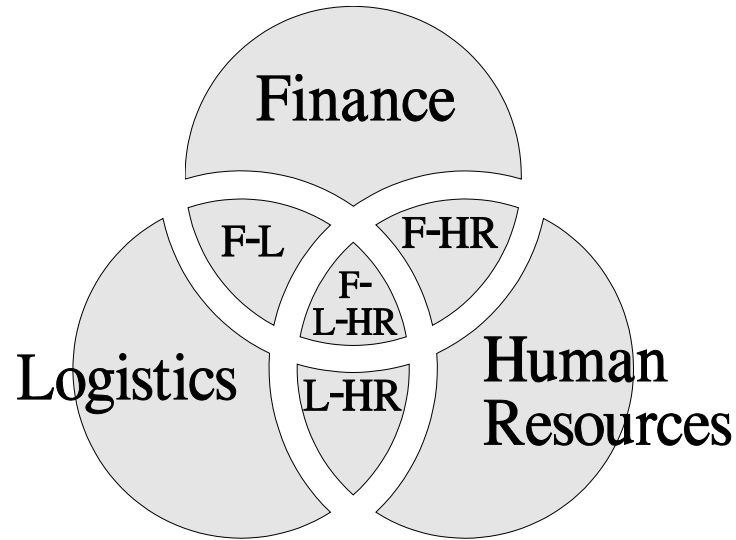
- Given each COI has 200 systems within each “domain.”
- The U.S. Army would then have 375 COIs.
- Given Integrated Data Exchange (IDE) approach means each COI has only 25 interfaces to standardize, NOT 600.
- COI interfaces cost is about \$8 million per year ( $25 * \$335K$ ), not \$200 million.

*Note: To eliminate COIs, or reduce their workload, eliminate systems*





### 3.4 Standardize Only at the “Shared Intersections.”

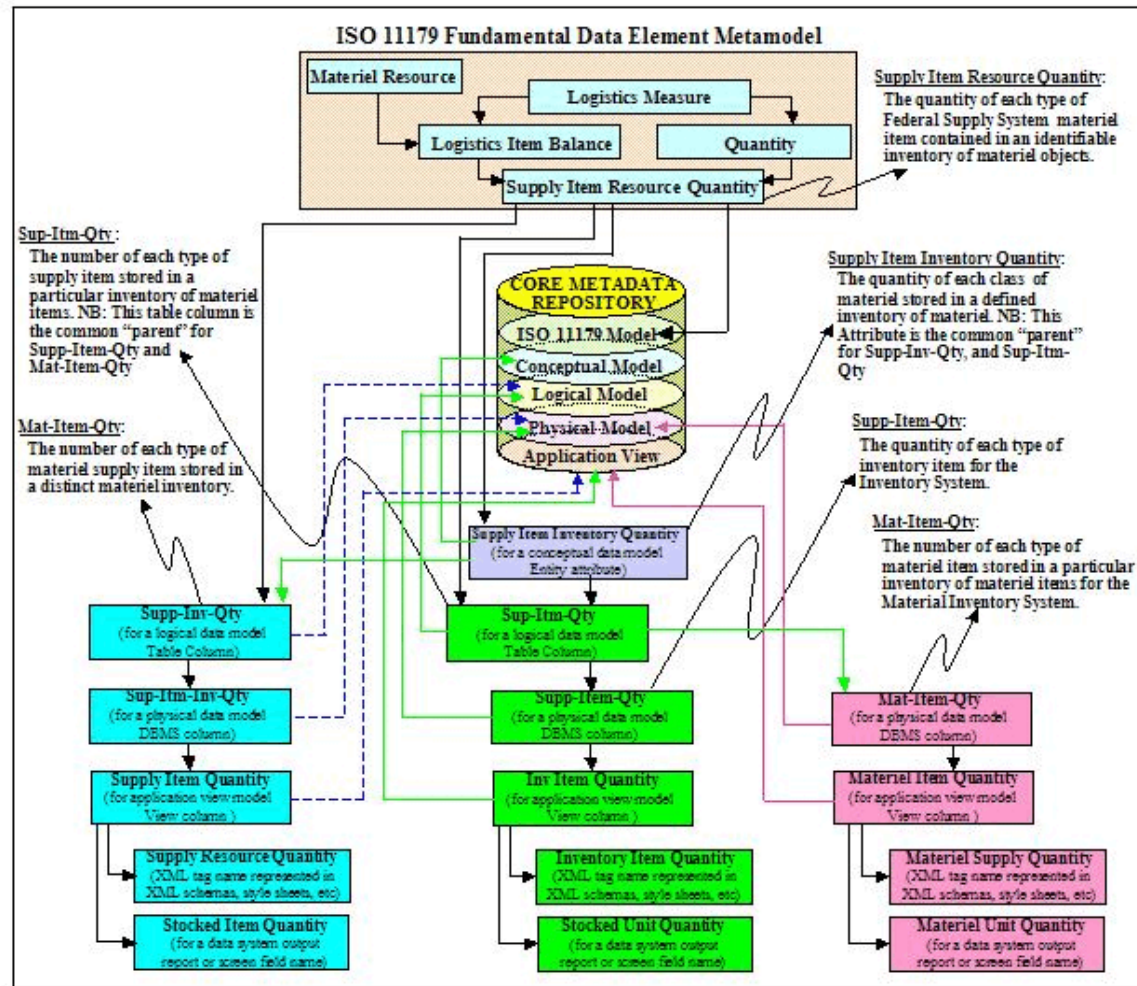


**If each database has 3000 columns, and only 15% overlaps, then**

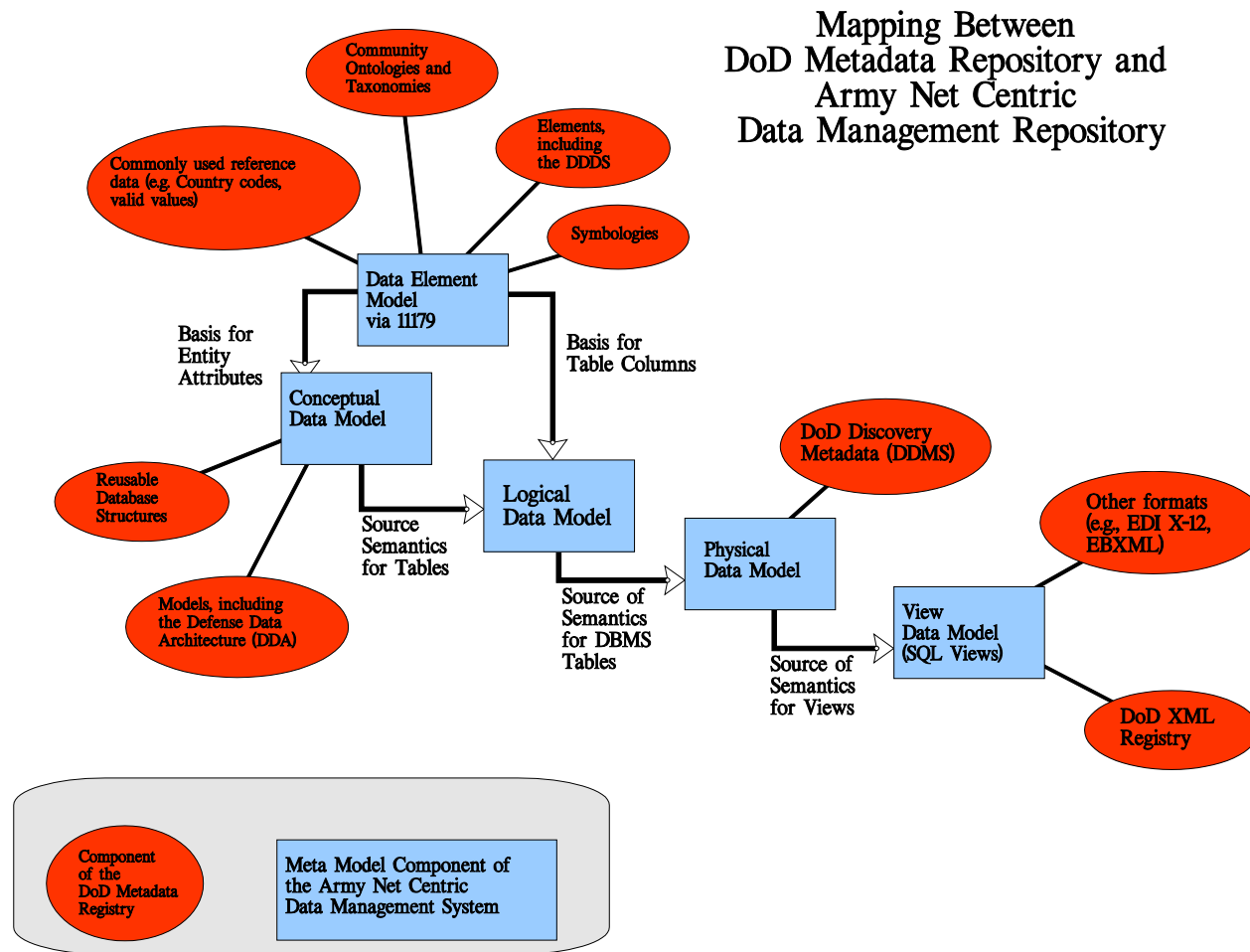
- F-L, F-HR, and L-HR are 450 columns each of shared data and semantics
- F-L-HR is only about 60 columns of shared data and semantics
- Standardization cost is only 15.6%: about 1.5 staff years vs 8.65 staff years
- Army standardization is thus reduced from  $(375 * 8.6)$  to  $(375 * 1.5)$



### 3.5 As Data Standardization Advances, Data Models & Metadata Converges

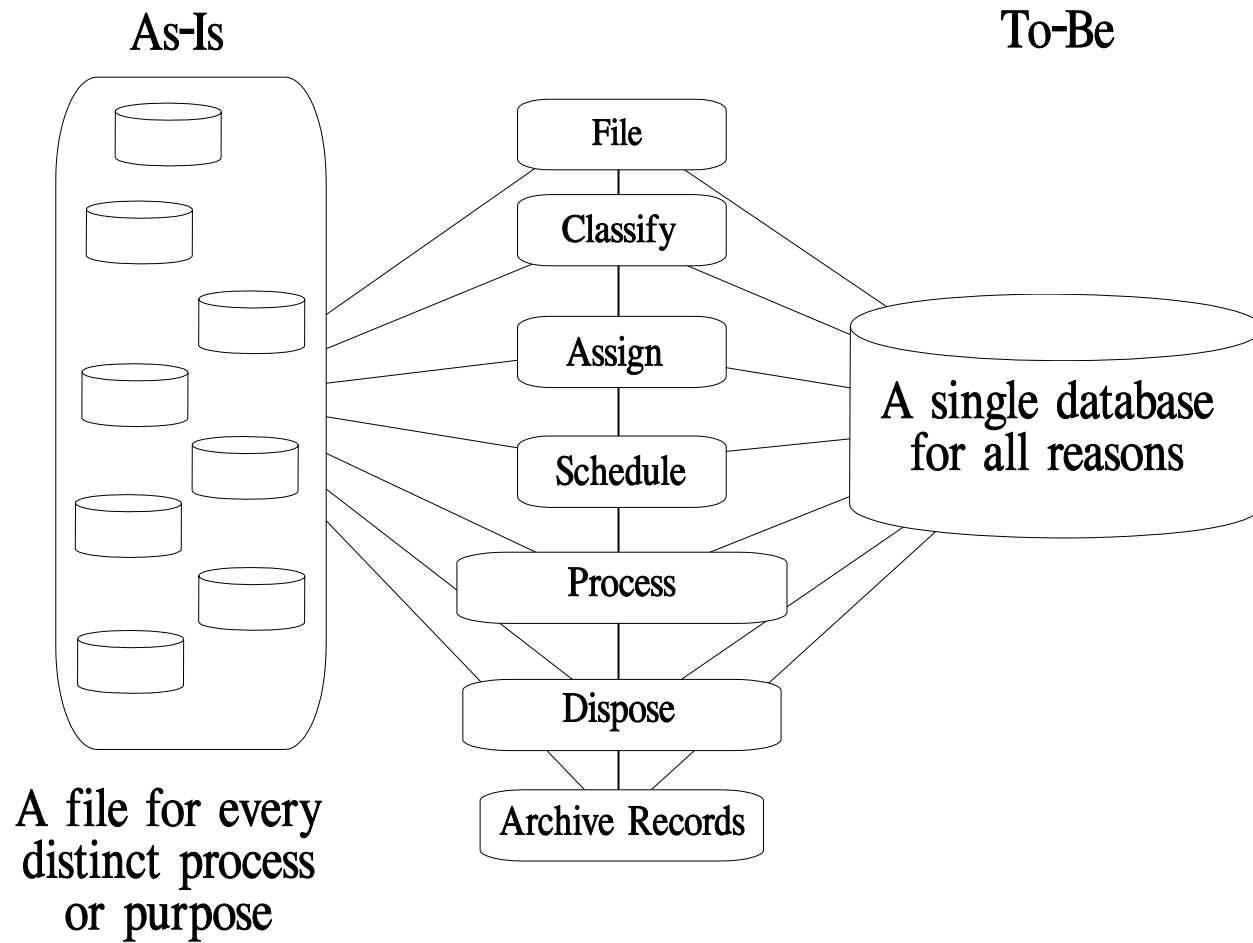


### 3.6 Build infrastructure to “automate” database design, and generate XML Schemas and Discovery Metadata.

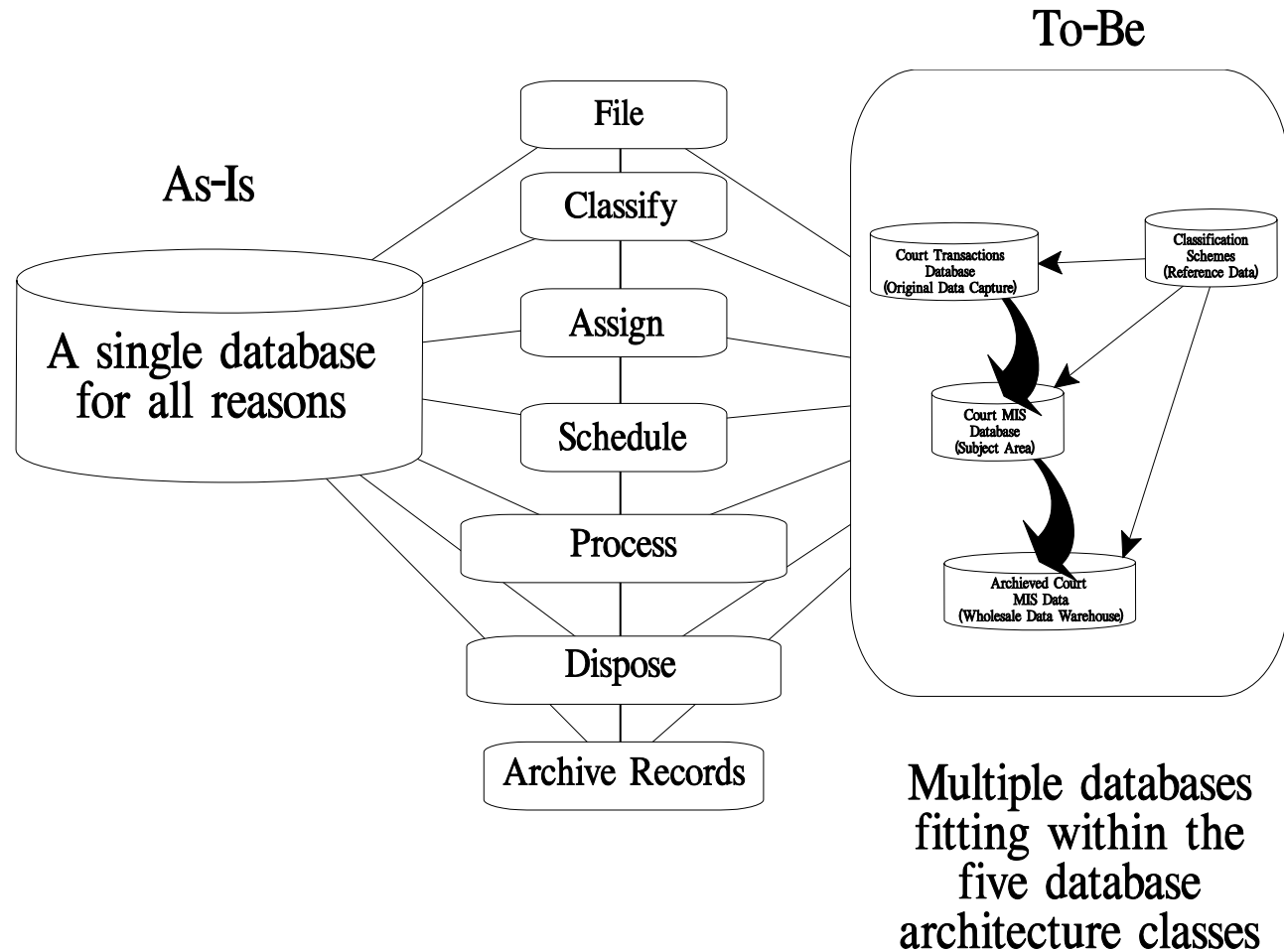




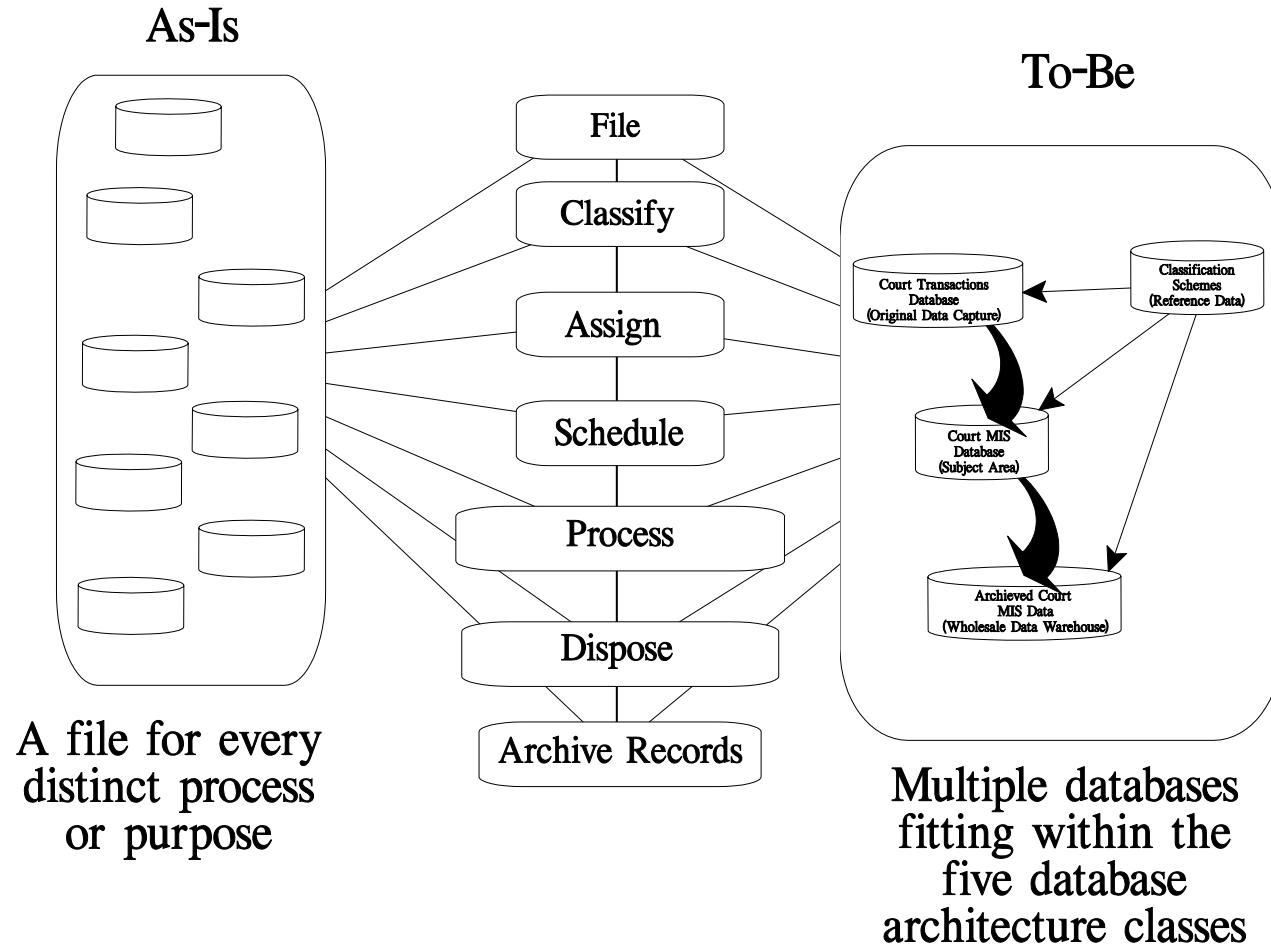
### 3.7.1 Consolidate Databases into Fewer and More Expansive



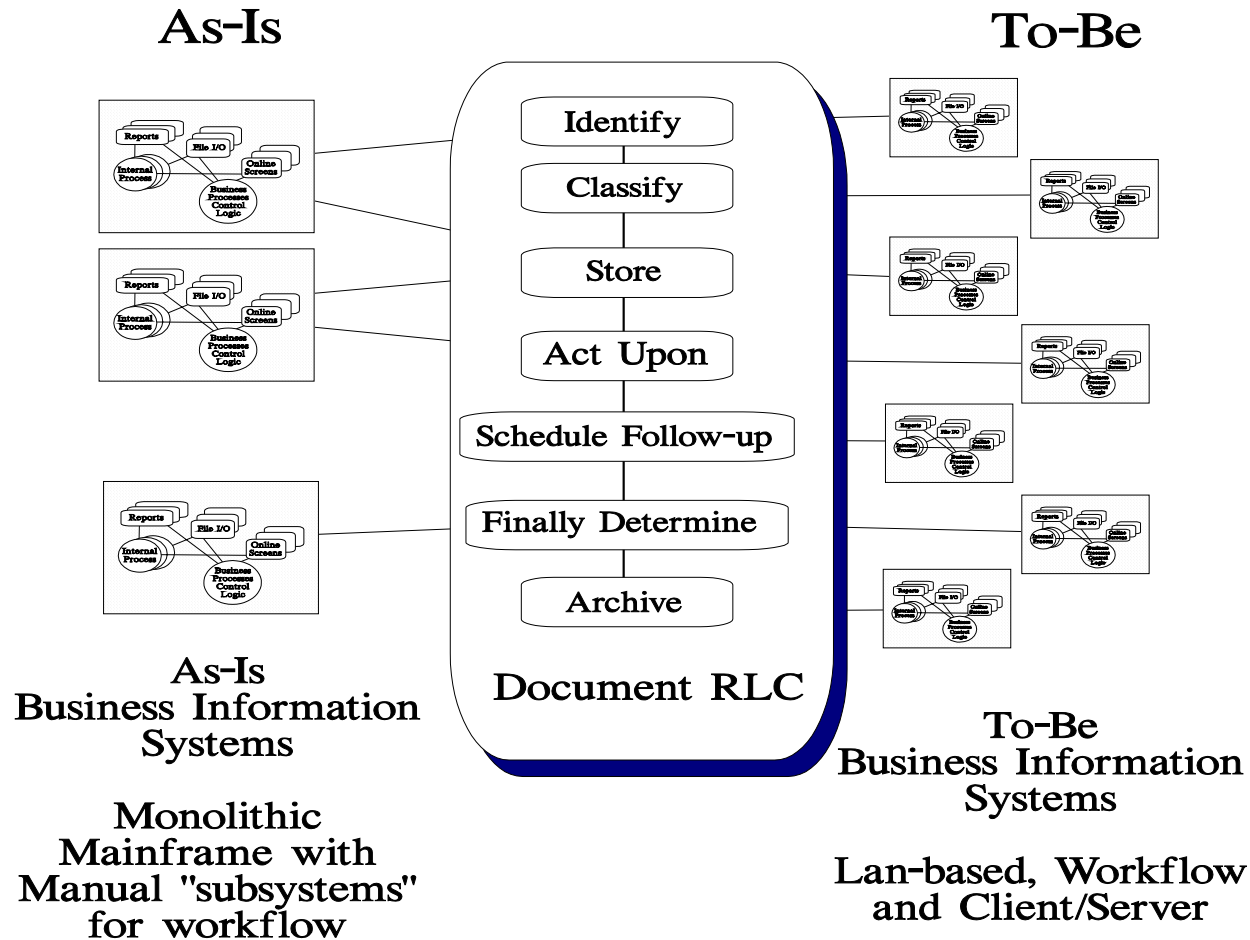
## Or, Single Database into Multi-tiered databases.



## Or Consolidate into a Multi-tiered database environment

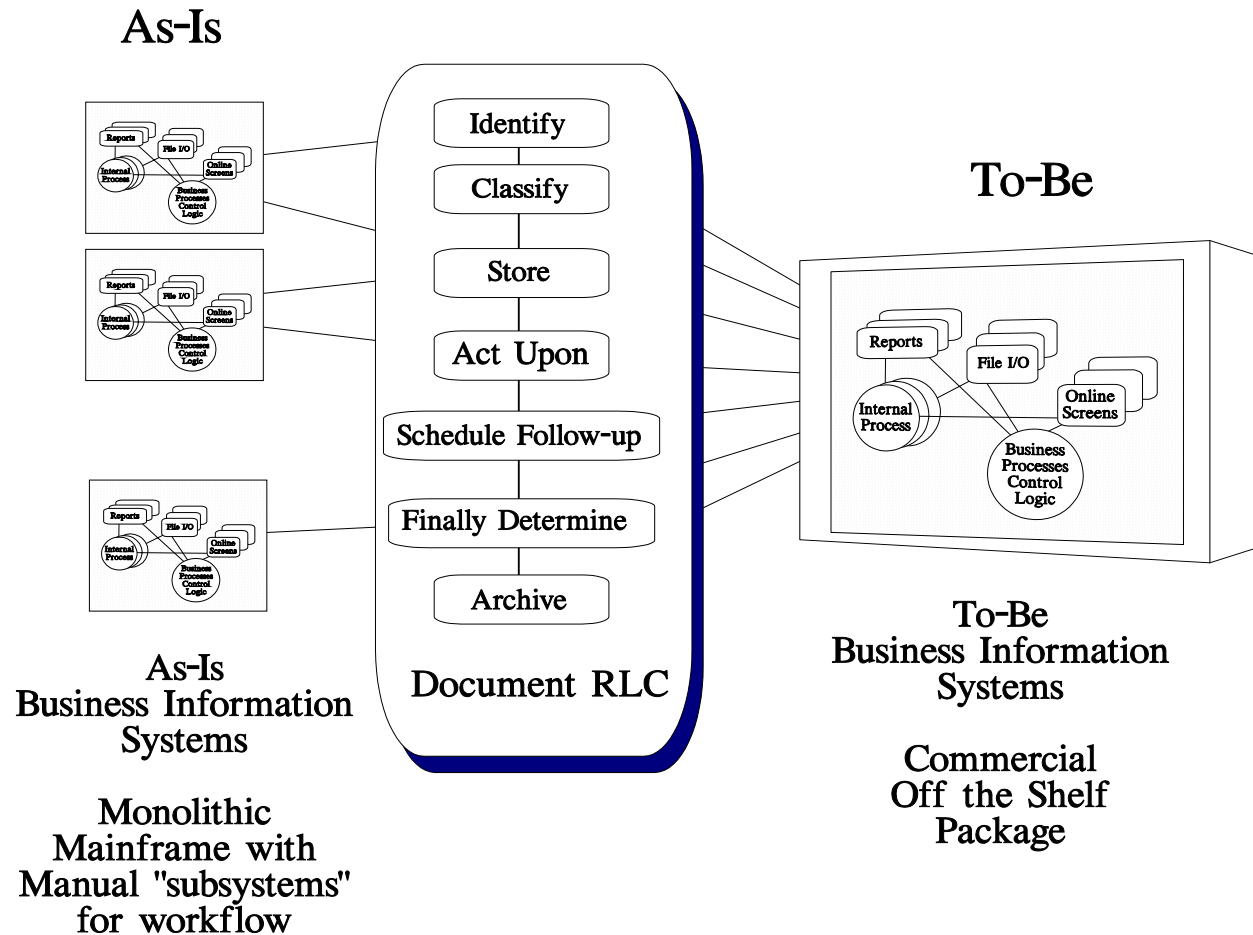


### 3.7.2 Re-cast Monolithic Systems into More Agile, Distributed Systems.

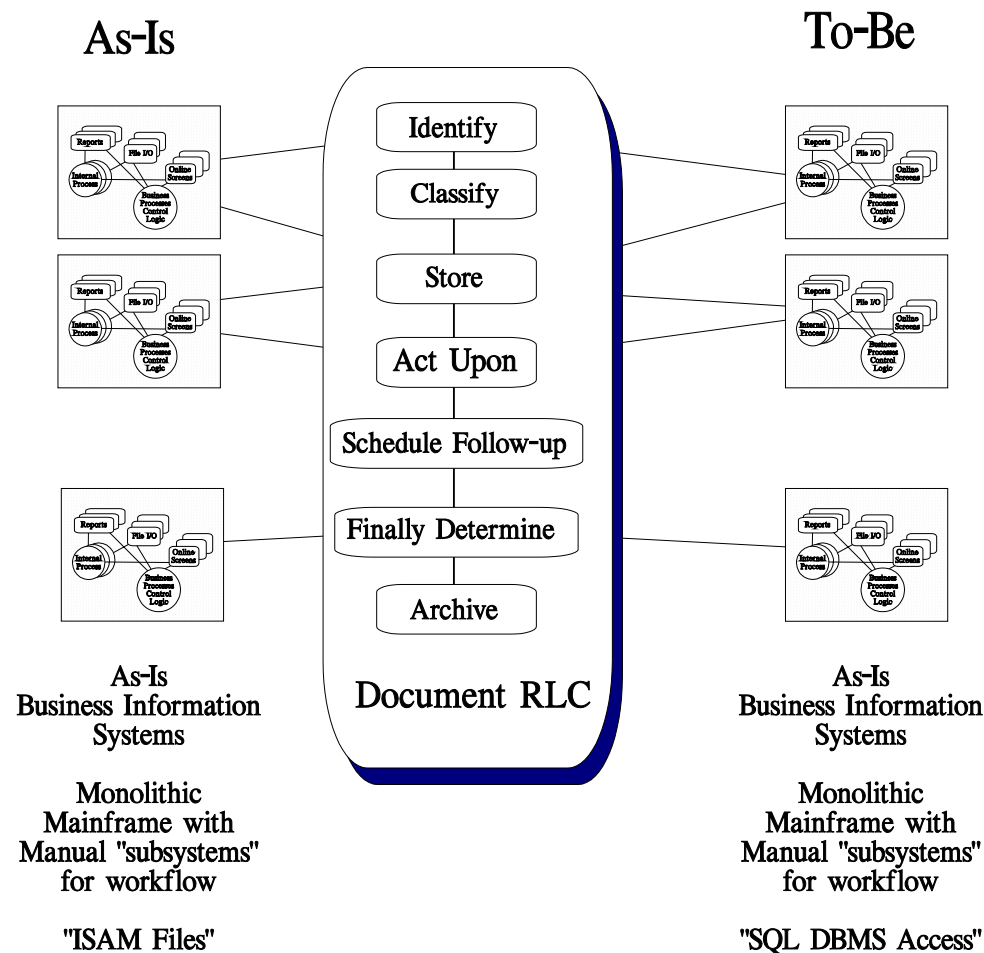




## Or, Monolithic systems into COTS ERPs



**Or, Change Standard access to SQL DBMS Access to broaden classes of access.**



## 4.0 Applicability to Prototypical System

### 4.1 Cost of Data Model Development

| Quantity of Tables<br>From estimate<br>of a<br>prototypical<br>database in<br>support of a<br>system | Average<br>Columns<br>per Table | Total<br>Columns | Process<br>Driven<br>Approach (2<br>hours per) | Data Driven<br>Approach (2<br>hours per<br>table, and<br>1/30th 11179<br>Data<br>Elements |  |
|--|---------------------------------|------------------|--|---|--|
| 400  | 15                              | 6,000            | 5.75 staff<br>years                            | 0.7 Staff<br>years  | <b>Cost Difference at<br/>\$100 per hour</b> |
| <b>Cost</b>  |                                 |                  | \$1,200,000                                    | \$120,000   | \$1,180,000 in favor of<br>data driven.      |



## 4.2 Cost of Prototypical System Software

| Quantity of Tables                       | Function Points Per Table | Cost Per Function Point | Type of Software | Total Cost    | Cost Difference (in favor of data driven) |
|--|---------------------------|-------------------------|------------------|---------------|---|
| 400 (If Data Driven)                     | 80                        | \$400                   | Information      | \$12,800,000  | \$43,008,000                              |
| 1744 (If Process Driven<br>(400 * 4.36)) |                           |                         |                  | \$55,808,000  |   |
| 400 (If Data Driven)                     | 80                        | \$1,000                 | Military         | \$32,000,000  | \$107,520,000                             |
| 1744 (If Process Driven<br>(400 * 4.36)) |                           |                         |                  | \$139,520,000 |   |

**Estimates exclude hardware, computing infrastructure, travel, testing, documentation, evolution and maintenance.**

