EWSolutions

Implementing a Successful Enterprise Data Quality Initiative

By David Marco

President

EWSolutions

EWSolutions' Background

EWSolutions is a Chicago-headquartered strategic partner and full life-cycle systems integrator providing both **award winning** strategic consulting and **full-service implementation services**. This combination affords our clients a full range of services for any size enterprise information management, managed meta data environment, and/or data warehouse/business intelligence initiative. Our notable client projects have been featured in the Chicago Tribune, Federal Computer Weekly, Crain's Chicago Business, and won the 2004 Intelligent Enterprise's RealWare award, 2007 Excellence in Information Integrity Award nomination and DM Review's 2005 World Class Solutions award.

Information Integrity Coalition
2007 Excellence in Information Integrity Award Nomination

Chicago Tribune

intelligent enterprise **REAL** Awards **TRANSFORM**
2004 WINNER
Best Business Intelligence Application Information Integration Client: Department of Defense

DM Review
2005
World Class Solutions Award Data Management

For more information on our Strategic Consulting Services, Implementation Services, or World-Class Training, call toll free at 866.EWS.1100, 866.397.1100, main number 630.920.0005 or email us at Info@EWSolutions.com

EWSolutions' Partial Client List

Arizona Supreme Court	Ford Motor Company	Neighborhood Health Plan
Bank of Montreal	GlaxoSmithKline	NORC
BankUnited	Harris Bank	Physicians Mutual Insurance
Basic American Foods	The Hartford	Pillsbury
Becton, Dickinson and Company	Harvard Pilgrim HealthCare	Quintiles
Blue Cross Blue Shield companies	Health Care Services Corporation	Sallie Mae
Branch Banking & Trust (BB&T)	Hewitt Associates	Schneider National
British Petroleum (BP)	HP (Hewlett-Packard)	Secretary of Defense/Logistics
California DMV	Information Resources Inc.	South Orange County Community College
College Board	International Paper	SunTrust Bank
Corning Cable Systems	Janus Mutual Funds	Target Corporation
Countrywide Financial	Johnson Controls	The Regence Group
Defense Logistics Agency (DLA)	Key Bank	Thomson Multimedia (RCA)
Delta Dental	LiquidNet	United Health Group
Department of Defense (DoD)	Loyola Medical Center	United States Air Force
Driehaus Capital Management	Manulife Financial	United States Navy
Eli Lilly and Company	Mayo Clinic	United States Transportation Command
Federal Aviation Administration	Microsoft	USAA
Federal Bureau of Investigation (FBI)	National City Bank	Wells Fargo
Fidelity Information Services	Nationwide	Wisconsin Department of Transportation
		Zurich Cantonal Bank

GSA Schedule
Contract GS-35F-0453M

SeaPort-e

For more information on our Strategic Consulting Services, Implementation Services, or World-Class Training, call toll free at 866.EWS.1100, 866.397.1100, main number 630.920.0005 or email us at Info@EWSolutions.com

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 3

Strategic Partner & Systems Integrator
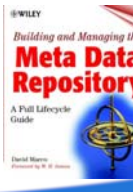Intelligent Business Intelligence[sm]

Professional Profile/Contact Information

Mr. Marco is an internationally recognized expert in the field of enterprise information management, data warehousing, Capability Maturity Model (CMM), business intelligence, and is the **world's foremost authority on meta data management**. Mr. Marco has authored several books including the widely acclaimed **"Universal Meta Data Models"** (Wiley, 2004) and the classic **"Building and Managing the Meta Data Repository: A Full Life-Cycle Guide"** (Wiley, 2000). These groundbreaking books have been broadly endorsed by many of the largest software companies in the industry and by several major magazines.

❑ Selected to the prestigious **2004 Crain's Chicago Business "Top 40 Under 40"**
❑ Crain's Chicago Business anointed him the "Melvil Dewey of Metadata"
❑ **2008 DAMA Data Management Hall of Fame (**Professional Achievement Award**)**
❑ Chairman of the Enterprise Information Management Institute (EIMInstitute.ORG)
❑ **2007 DePaul University named** him one of their **"Top 14 Alumni Under 40"**
❑ Presented hundreds of keynotes/seminars across four continents
❑ Published hundreds of articles on information technology
❑ Author of several best selling information technology books
❑ Taught at the **University of Chicago** and **DePaul University**
❑ Judged dozens of various industry awards in meta data management and data warehousing
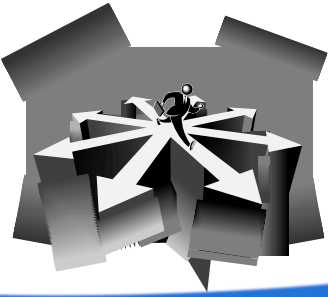
Email: DMarco@EWSolutions.com

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 4

Strategic Partner & Systems Integrator
Intelligent Business Intelligence[sm]

Agenda

- ❑ The State of Data Quality
- ❑ Data Quality Challenges
- ❑ Data Quality Foundations
- ❑ Data Quality Solutions
- ❑ Real-World Data Quality Case Studies

The State of Data Quality

The State of Data Quality

❑ Most organizations data quality can be best described as abysmal

❑ Large bank's Executive VP stated:("The numbers (data in their data warehouse) are the best that we have,)and(as long as they are calculated the same way every time,) (they give us a directionally correct view of the business.")

The State of Data Quality

❑ When data is presented to a business analyst or executive typically the person does not know:
  ➢ if the data accurate?
  ➢ what percentage of the data is not accurate or in doubt
  ➢ the origin of the data
  ➢ the meaning of the data
  ➢ where the data came from

❑ Our information technology (IT) system's architectures are not conducive to data quality

What Does One Process Look Like for a Large Company?

Fortune 50 – One Process

Maybe this Image Clears When you Look at One Application?

Islands of Data
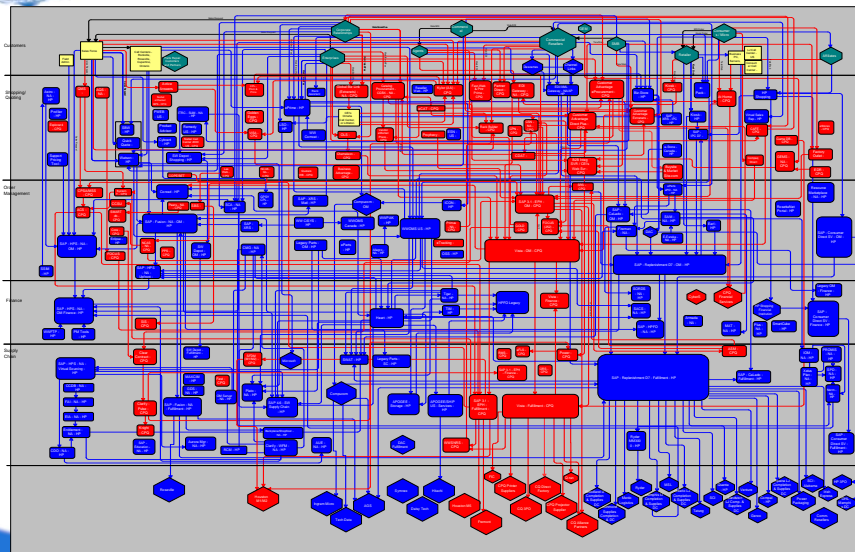
How Did We Get Here

- ❑ Most companies did not plan their infrastructure…rather it just grew over time
- ❑ Previously companies focused on lines-of-business, as opposed to an enterprise view

Why is Data Quality Growing as an Issue?

- ❑ Enterprise initiatives (e.g. data warehousing, customer relationship management, supply chain, etc.) are most significantly impacted by data quality issues
- ❑ Corporations are moving from a decentralized structure to a centralized structure
- ❑ Government regulations require data quality (SOX, 21 CFR Part 11, BASIL II, HIPPA, DoDAF, ISO 11179, etc.)

Data Quality Solutions

Data Quality Solutions

❑ *Data Quality Falls Under the Larger Enterprise Information Management Umbrella*

❑ *Meta Data Management is the Technical Enabler of Data Quality*

❑ *Data Stewardship defines the Business Processes of Data Quality*

Data Quality
&
Enterprise Information
Management

Data Quality & EIM

❑ Many companies rush to implement data quality "silo" solutions without realizing that they need to understand the "larger picture"

❑ Data quality falls under the larger Enterprise Information Management (EIM) umbrella

What is EIM?

❑ **Enterprise Information Management (EIM):** The systematic processes and governance procedures for applications, processes, data, and technology at a holistic enterprise perspective

❑ The purpose of enterprise information management is to bring enterprise order, purpose, structure, efficiency, and performance to applications, processes, data, meta data and technology

❑ EIM is not a single technology or component, but a coordinated framework of disciplines for managing data, meta data and information assets throughout the organization

❑ *Data Does Not Manage Itself!!*

Data Quality & EIM

❑ There are three foundational elements that span each area of EIM

➢ Meta Data Management

➢ Data Governance

➢ Data Management

❑ No matter what focus area of EIM you are targeting you will need to address each of these elements

Data Quality & EIM

Enterprise
Information Management

Data Management is the foundation for all of the other EIM focus areas.

| Process Management | Data Architecture | **Data Quality** | IT Portfolio Management | Master Data Management | Information Delivery | Information Security |

Data Management

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 21

Strategic Partner & Systems Integrator
Intelligent Business Intelligence^{sm}

Meta Data Management

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 22

Strategic Partner & Systems Integrator
Intelligent Business Intelligence^{sm}

Meta Data vs. Data

- **Meta Data:** Meta data contains the knowledge that a **1)** field is called "Customer_Name", is 40 characters in length, and exists in systems A, B, and C; **2)** that our company has 3 systems which contain customer master data. These systems are…

- **Data:** Data would be a specific instance of "Customer_Name" equaling "John Doe"

- **Information:** Data that is meaningful to a business user. They understand it and they know what to do with it

www.EWSolutions.com
Strategic Partner & Systems Integrator
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 23
Intelligent Business Intelligence℠

EIM Fundamentals

Information = Data + Meta Data
(content)   *(context)*

www.EWSolutions.com
Strategic Partner & Systems Integrator
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 24
Intelligent Business Intelligence℠

Integrating DQ into IT

2007 Monthly Global Sales Report   **April 7, 2008**

Month	Product Category	Sales $ (in thousands) U.S	Sales $ (in thousands International	Sales $ (in thousands Total
December	TV	22,10	10,200	32,301
	VCR	11,190	4,300	15,490
	Cellular Phone	12,190	7,193	19,383
	Digital	4,002	1,301	5,303
	Miscellaneous	1,209	870	2,079
November	TV	42,000	22,200	64,200
	VCR	21,190	9,878	31,068
	Cellular Phone	28,193	12,193	40,386
	Digital	8,901	901	11,802
	Miscellaneous	2,730		4,260
October	TV	70,100	32,	103,050
	VCR	31,900	14,8	46,778
	Cellular Phone	41,700	17,550	59,250
	Digital	20,000	4,100	24,100
	Miscellaneous	4,850	2,850	7,700

Meta Data

"Sales $ U.S." is comprised of aggregated sales revenues from the United States, Canada, and Mexico, but does not subtract sales dollars from returned orders

Strategic Partner & Systems Integrator
Intelligent Business Intelligence[sm]

Integrating DQ into IT

2007 Monthly Global Sales Report   **April 7, 2008**

Month	Product Category	Sales $ (in thousands) U.S	Sales $ (in thousands International	Sales $ (in thousands Total
December	TV	22,101	10,200	32,301
	VCR	11,190	4,300	15,490
	Cellular Phone	12,190	7,193	19,383
	Digital	4,002	1,301	5,303
	Miscellaneous	1,209	870	2,079
November	TV	42,000	22,200	64,200
	VCR	21,190	9,878	31,068
	Cellular Phone	28,193	12,193	40,386
	Digital	8,901	2,901	11,802
	Miscellaneous	2,730	1,530	4,260
October	TV	70,100	32,950	103,050
	VCR	31,900	878	46,778
	Cellular Phone	41,700		59,250
	Digital	20,000		24,100
	Miscellaneous	4,850	2,8	7,700

Meta Data

Data Quality Tracking Statistics
8.4% of the dollar values were not loaded
1.7% of the records were not loaded

Strategic Partner & Systems Integrator
Intelligent Business Intelligence[sm]

Does Understanding Your Data Lead To Better Decisions?

NASA Example – Mars Orbiter

- ❑ December 11, 1998 the Mars Climate Orbiter was launched
- ❑ Engineers calculated rocket firing using feet-per-second, the orbiter was programmed in meters-per-second (metric system) of thrust
- ❑ The difference was 4.4 feet per second
- ❑ "Each time there was a burn (rocket firing) the error built up," said Art Stephenson, Director of the Marshall Spaceflight Center and Head of the NASA Investigation Team
- ❑ "We entered the Mars atmosphere at a much lower altitude (than planned)," said Ed Weiler, NASA's chief scientist. "It (the spacecraft) either burned up in the Martian atmosphere or sped out (into space). We're not sure which happened."
- ❑ The cost of this mission was $250 - $300 million

Source: Associated Press, Paul Recer, http://www.anomalous-images.com/news/news537.html

Integrating DQ into IT

Campaign Analysis Query		FirstRateMotors			January 20, 2008
Campaign Dates	**Campaign Name**	**Campaign Description**	**Auto Sales**	**Auto Type**	
01/01/2003 - 03/31/2003	Buy Now, Pay Later	The "Buy Now, Pay Later" campaign was a nation-wide campaign. This campaign offered customers the option not to begin payments for a new car purchase until January 1, 2004. This offer is only applicable for those customers with better than standard credit ratings.	7,045	ECON – Smaller, cheaper, economy car line	
			9,500	MID – Mid-size, sedan automobiles	
			22,010	SPTS – Sports/performance cars	
			14,700	SUV – Sport Utility Vehicles	
		Grand Total	**53,255**		
04/01/2003 - 06/30/2003	Buy One, Get A Free Scooter	The "Buy One, Get A Free Scooter" campaign was a nation-wide campaign. This campaign offered customers a free scooter for the purchase of a new car. Dealer can only discount cars, up to $1,000	5,205	ECON – Smaller, cheaper, economy car line	
			7,250	MID – Mid-size, sedan automobiles	
			17,888	SPTS – Sports/performance cars	
			10,900	SUV – Sport Utility Vehicles	
		Grand Total	**41,243**		
07/01/2003 - 09/30/2003	$1... Tra...	...campaign was a ...paign offered customers ...mum trade-in, regardless of the condition... -in vehicle, only for those customers p... new car.	6,102	ECON – Smaller, cheaper, economy car line	
			8,330	MID – Mid-size, sedan automobiles	
			19,750	SPTS – Sports/performance cars	
			12,400	SUV – Sport Utility Vehicles	
		Grand Total	**46,582**		
10/01/2003 - 12/31/2003	0 down and 0% Interest	The "0 down and 0% Interest" campaign was a nation-wide campaign. This campaign offered customers standard car discounts, with 0 ... s and 0% interest for only those customers with sta... d or better credit ratings.	6,700	ECON – Smaller, cheaper, economy car line	
			8,925	MID – Mid-size, sedan automobiles	
			20,820	SPTS – Sports/performance cars	
			13,220	SUV – Sport Utility Vehicles	
		Grand Total	**49,665**		

Data Quality Statistics
2.3% of the records were not loaded in the data warehouse batch runs.
Skew percentage equals 2.3%

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 29
Strategic Partner & Systems Integrator
Intelligent Business Intelligence℠

Data Quality and Meta Data Management

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 30
Strategic Partner & Systems Integrator
Intelligent Business Intelligence℠

Data Quality and Meta Data Management

❑ Good data quality professionals understand the that meta data management is their *key technical enabler*

❑ Where do business rules for data quality go?
  ➢ a spreadsheet?
  ➢ a document?
  ➢ maybe we can just commit them to memory?
  ➢ in a Managed Meta Data Environment (MME)

❑ What is a data quality rule? Its meta data!

❑ Let's look at some MME implementations that have strong focuses on data quality

Data Stewardship

Data Stewardship

- **Data Stewardship:** The process of having data stewards work with the data and meta data of an organization to ensure its quality, accuracy, formats, domain values, and that it is properly defined and understood across the enterprise
- **Data Steward:** A person(s) responsible for working with the data and meta data. There are different types of data stewards
- The data steward acts as the conduit between IT and the business. The data steward (often not just one person, but a collection of people) aligns the IT systems (both decision support and operational) with the business' requirements. The data steward has the challenge of guaranteeing that one of the corporation's most critical assets--its data--is used to its fullest capacity

Data Stewardship

- Data is one of the most important assets in a corporation
- Data has value when it is delivered timely, properly formatted, concise, accurate and **understood**
- Business ownership of the data and active participation are critical
- The role of the data steward has grown considerably over the years

# Data Governance Organization

❑ Every organization forms their data governance organization a little differently

❑ Some have a more or less complex organization

❑ What is critical is that the organization:

- ➢ is actively using the MME
- ➢ has clear lines of communication
- ➢ has a defined and well understood decision making process
- ➢ well defined feedback loop

Data Governance Organization
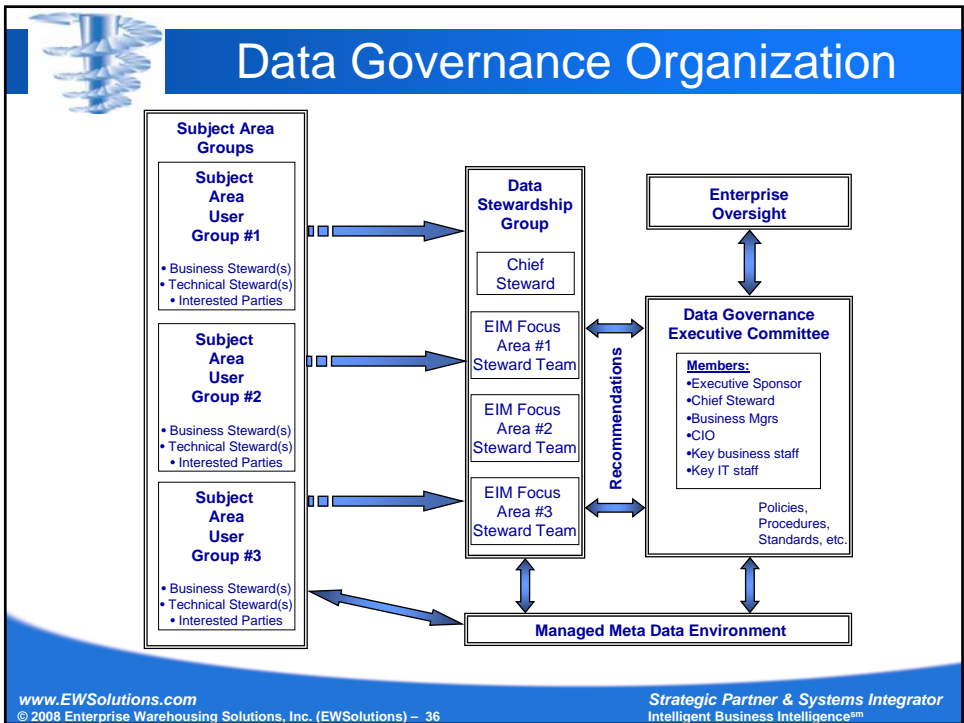
Real-World MME Implementations

Allstate Corporation

The following slides are derived from the book "Universal Meta Data Models", David Marco & Michael Jennings, Wiley 2004

Allstate Corporation

- ❑ The nation's largest publicly held personal lines insurer
- ❑ Provides insurance products to more than 16 million households and has approximately 12,300 exclusive agents and financial specialists in the United States and Canada
- ❑ Multi-channel organization: customers can access products and services through Allstate agents, or in select states at allstate.com and 1-800-Allstate®
- ❑ Allstate Financial Group includes the businesses that provide life and supplemental insurance, retirement, banking, and investment products through distribution channels that include Allstate agents, independent agents, and banks, and securities firms
- ❑ *The MME was part of a larger Data Asset Management effort*

www.EWSolutions.com
Strategic Partner & Systems Integrator
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 39
Intelligent Business Intelligence℠

Allstate Corporation

Challenges/Opportunity

- ❑ Early 1990s Allstate, like most large corporations, found itself challenged with managing disparate systems to satisfy its IT needs
- ❑ Needed applications to talk across platforms
- ❑ Systems that have different coding schemes for common codes and mismatches in field types and sizes cannot interchange data easily
- ❑ Dramatically reduce the number of point-to-point interfaces
- ❑ Needed a precise understanding and knowledge of the data that the analysts of the data warehouse would utilize

www.EWSolutions.com
Strategic Partner & Systems Integrator
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 40
Intelligent Business Intelligence℠

Allstate Corporation

Challenges/Opportunity

❑ Need to Manage Code Data
- ➢ Very "code-driven" environment because every state has its own regulatory statutes to which Allstate must adhere
- ➢ Highly time-consuming and difficult to integrate systems because projects would take longer to deliver and be too costly
- ➢ Sound meta data management techniques was expected to reduce or eliminate IT rework, speed up projects, and lower their overall costs

❑ Enable Data Warehousing Applications
- ➢ Data warehousing made data quality more important than ever
- ➢ Previously subject matter experts in an application area had to know what the data represented, what it actually meant, and how to use it
- ➢ Now this data was going to be presented to actual end users, and Allstate cannot afford to have user "interpretations" of what the data *may* mean be used to make "live" business decisions.
- ➢ A strong, centralized data management environment would be the basis for consistent data driving high-quality decisions by end users
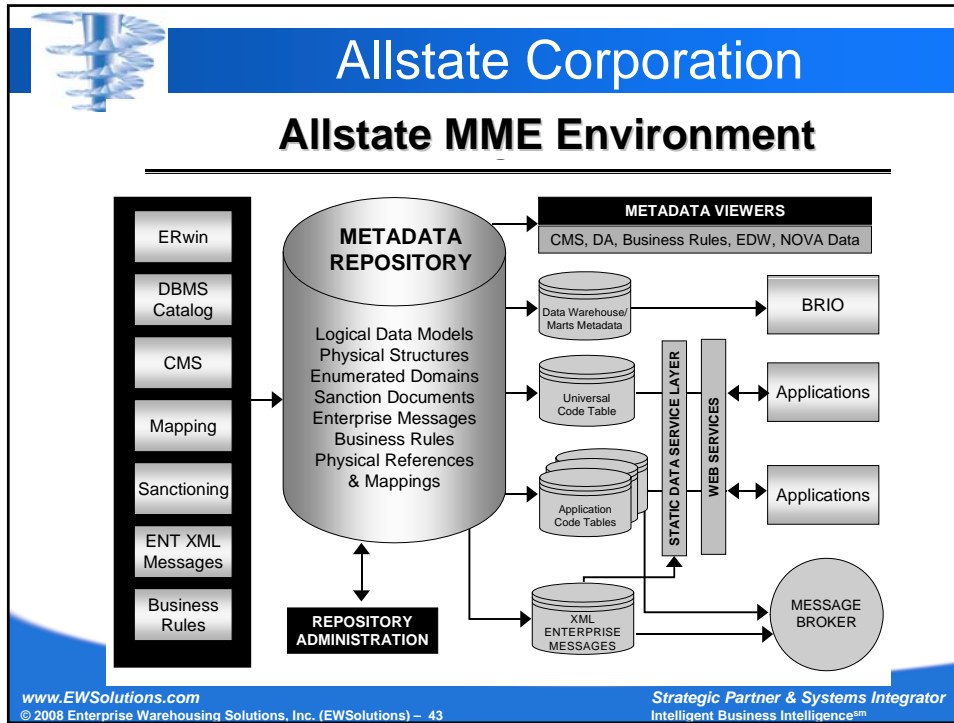
Allstate Corporation

MME Overview

❑ Custom built MME

❑ Initially focused on the management of the codes that permeated their systems

❑ A portion of the MME (Codes Management System (CMS)) allows for the identification of enumerated domains (those for which a set list of values can be listed) and define the various coding schemes that were found in the different applications and the associated business values

❑ CMS allowed Allstate's Codes Analysts group to do their job more effectively

❑ Codes analysts, along with a group of data administrators, then became the nucleus of the Enterprise Data Management group

❑ Kept an enterprise perspective by documenting each unique domain they encountered and storing it in the MME

❑ As they worked with subsequent projects, they were able to see where the same data had been encoded differently between applications

Allstate Corporation

Allstate MME Environment

(Diagram: ERwin, DBMS Catalog, CMS, Mapping, Sanctioning, ENT XML Messages, Business Rules feeding into METADATA REPOSITORY containing Logical Data Models, Physical Structures, Enumerated Domains, Sanction Documents, Enterprise Messages, Business Rules, Physical References & Mappings. Repository connects to Metadata Viewers (CMS, DA, Business Rules, EDW, NOVA Data), Data Warehouse/Marts Metadata → BRIO, Universal Code Table → Applications, Application Code Tables → Applications through Static Data Service Layer and Web Services, XML Enterprise Messages → Message Broker, and Repository Administration.)

Allstate Corporation

Data Stewardship

- January 1997 formed a Data Stewardship Council (DSC)
- DSC is a cross-business-unit team focused on the business aspects of managing data as a valued enterprise asset
- DSC is a part-time, virtual team of Allstate employees who have strong business knowledge, vision, and the ability to look horizontally across the enterprise
- Data stewards are focused on addressing the business issues behind key data resource management objectives: managing data redundancy, implementing data shareability and standardization, and managing and improving data integrity
- Stewards follow several basic principles for managing data resources of any type; these include the following:
  - Requirements for the resource must be anticipated and fulfilled proactively
  - Allstate cannot afford an infinite amount of the data resource; therefore, the amount must be optimized
  - Data resource should be shared and leveraged in as many ways as possible, in order to maximize its value while diminishing its overall costs
  - Data resource must be carefully managed to ensure that its use in the business is prudent, efficient, effective, and secure

Department of Defense

intelligent enterprise REAL Awards TRANSFORM
2004 WINNER
Best Business Intelligence Application
Information Integration

How Large is the DoD Supply Chain?

- ❑ $480 Billion Dollars Budget
- ❑ $80 Billion Supply Chain
- ❑ 3 Million People
- ❑ 8 Million Parts
- ❑ Global – 150 Countries
- ❑ Dynamic Supply Chain

intelligent enterprise REAL Awards TRANSFORM
2004 WINNER
EWSolutions
Best Business Intelligence Application
Information Integration

What's the Challenge

- Can't Account for Several Billion Dollars a Year
- Have Data Quality Issues
- Fix or Refine a Process and Break Several Other Processes
- Never Planned the Enterprise
- The Enterprise "Just Grew"

DoD Enterprise Architecture is Based on:

Support rapid, agile deployment, employment, sustainment and reset/reconstitution Total Force,

within acceptable risk, across the full spectrum of operations in a cost-effective manner (ROI)

Enterprise Integrated Data Environment (EIDE)

Provide an enhanced environment that enables the DoD Logistics Enterprise to execute practices, processes, applications and decision support tools to achieve logistics interoperability and allow for information exchange within and between internal and external DoD business partners.

- Non-system dependent transactions
- Consolidation and reuse of Interfaces
- Data integration/sharing
- Leverage Modernization Efforts
- Data Standards not Standard Data

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 49
Strategic Partner & Systems Integrator
Intelligent Business Intelligence[sm]

DoD Enterprise Logistics

DoD ENTERPRISE LOGISTICS
Operational Activity to System Function Traceability Matrix (SV-5)

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 50
Strategic Partner & Systems Integrator
Intelligent Business Intelligence[sm]

Canonical Model and Legacy Information within the EIDE

Business Process
&
UID

Canonical Model

NSN	UID	Navy Part #	Supplier Part #
1615011975441	25.77272.34	1615011975441AH	114R2080-9
1615013005539	45.97499.43	1615013005539AH	540-011-101-129

Distributed Meta Data Architecture (DMDA)

Applying Compliance Architecture and Programs The Toolkit"

Architecture and Program Compliance
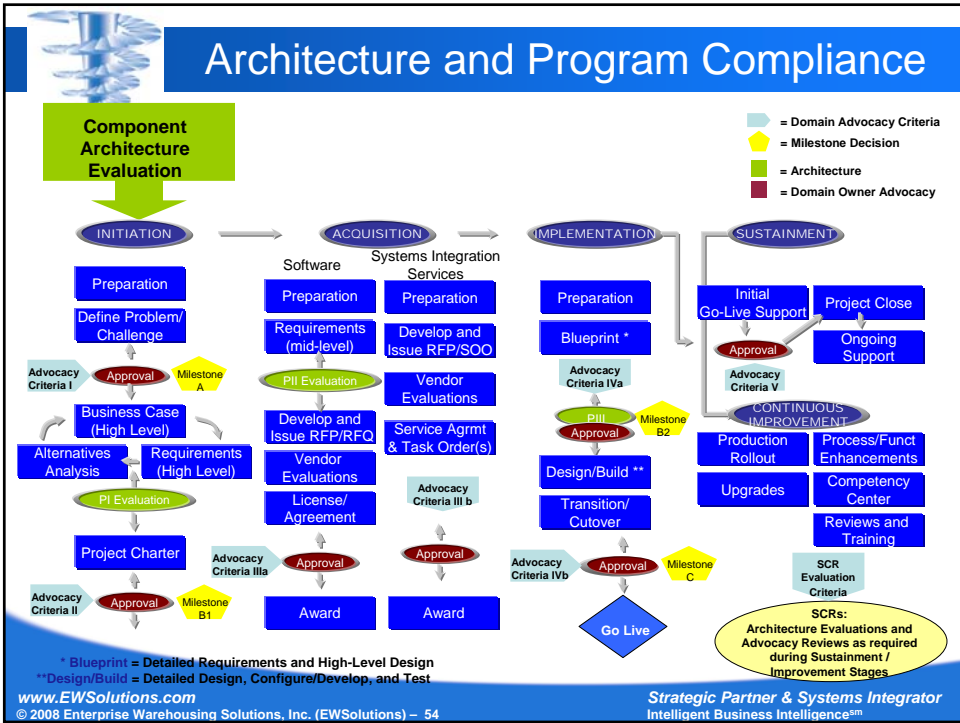
Summary

- ❏ Plan Your Enterprise Around the "To Be Business Processes" not your "As Is" (80-20 Rule)
- ❏ Data Strategy is the Key to Data Interoperability Across the Enterprise
- ❏ No Meta Data Management, No Data Strategy
- ❏ Compliance is an "On-Going Process"
- ❏ Your Business Processes Should Not Be Unique

intelligent enterprise® **REAL** Awards **TRANSFORM**
2004 WINNER
EWSolutions
Best Business Intelligence Application
Information Integration

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 55

Strategic Partner & Systems Integrator
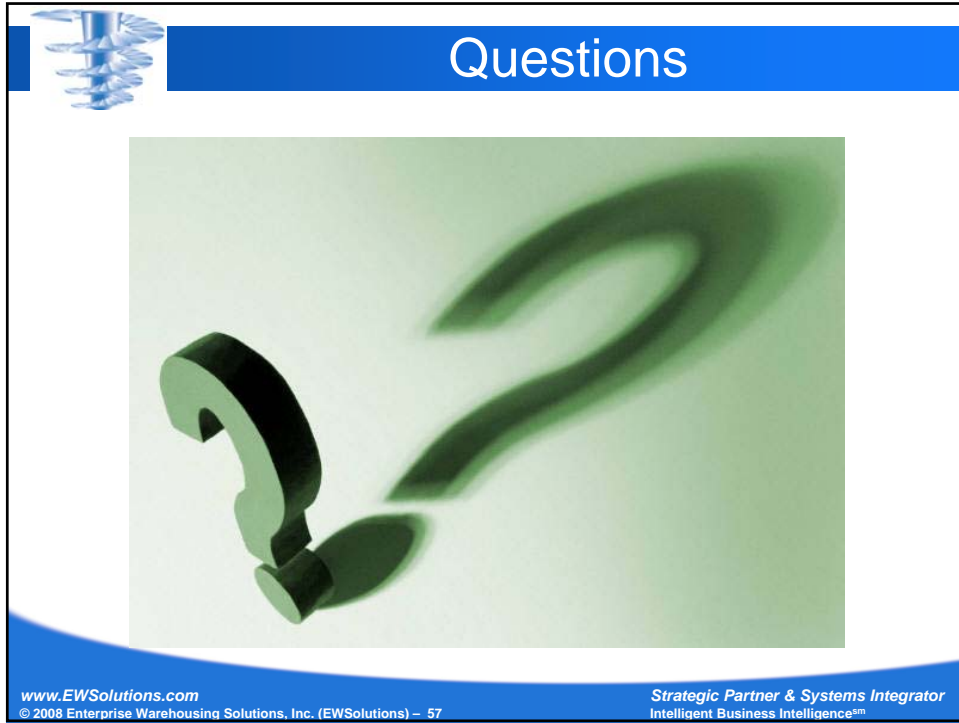Intelligent Business Intelligence℠

Don't Limit Yourself

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 56

Strategic Partner & Systems Integrator
Intelligent Business Intelligence℠